

From: Ramirez, Delia
Sent: Friday, February 22, 2002 11:09 AM
To: STIC-Biotech/ChemLib
Subject: case 09/819136

Point of Contact
Arlene Smith
Technical Information Specialist
CM1 6A01
Tel: 308-3278

Hi,

I would like to request the following searches (Conklin et al.):

1. a standard search of seq id 2 in the protein databases (commercial and interference)
2. an oligo search of seq id 2 in the protein databases (commercial)
3. a standard search of seq id 2 33-75, 93-157, 39-157, 203-286, 299-351, 412-548, 299-409, 33-548 in the protein databases (commercial and interference).
4. an oligo search of seq id 2 117-122, 525-530, 283-288, 50-55 (commercial)

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner - Art Unit 1652
USPTO
1911 S. Clark Street, Crystal Mall 1, 10D04, Mail room 10C01
Arlington, VA 22202
(703) 306-0288
delia.ramirez@uspto.gov

CRTG

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 2/25/02
Date Completed: 2/27/02
Searcher Prep/Review: 15
Clerical: _____
Online time: 15

TYPE OF SEARCH:
NA Sequences: 14
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2002, 23:58:55 ; Search time 76.6 Seconds
(without alignments)
529.924 Million cell updates/sec

Title: US-09-819-136-2
Perfect score: 3016
Sequence: 1 MPALRRLPLLLLRITSGA.....KKILELLEKQACELNRFQD 548

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	840	27.9	216	22	AAAB8333
2	358	11.9	59	21	AAVB4381
3	222.5	7.4	560	15	AAAR62523
4	213	7.1	143	20	AAV08609
5	213	7.1	147	17	AAAR92337
6	213	7.1	147	18	AAAR25928
7	213	7.1	352	9	AAAP81110
8	213	7.1	352	22	AAAU14343
9	213	7.1	352	22	AAU14579
10	213	7.1	352	22	AAE05095
11	213	7.1	366	21	AAAB43730

12	212	7.0	147	19	AAAR65222	rUTI protein sero I
13	210	7.0	124	18	AAAR59322	New protease inhib
14	210	7.0	144	18	AAAR25936	SUC2-Ep1-d21-RPDS
15	210	7.0	145	18	AAAR25935	Novel protease inh
16	210	7.0	165	18	AAAR25937	Elastase inhibitor
17	209	6.9	124	17	AAAR92233	Human mature urina
18	209	6.9	124	18	AAAR25933	New protease inhib
19	209	6.9	144	18	AAAR25938	SUC2-Ep7-d21-RPDS
20	209	6.9	145	17	AAAR22235	Human urinary tryp
21	207	6.9	122	20	AAAR08611	Inter-alpha-trypsi
22	195.5	6.5	122	13	AAAR24110	Trypsinogen structu
23	195.5	6.5	304	16	AAAR83899	Human lipoprotein-
24	194.5	6.4	256	12	AAAR11169	Ser(Asp1-Rth255)-E
25	194.5	6.4	261	12	AAAR11171	Ser(Asp1-Glu245)-
26	194.5	6.4	262	12	AAAR11172	Ser(Asp1-Ser248)-
27	194.5	6.4	265	12	AAAR11170	Ser(Asp1-Ile253)-
28	194.5	6.4	276	17	AAAR92265	TFPI mutein, K36R.
29	194.5	6.4	276	17	AAAR92012	Recombinant non-gl
30	194.5	6.4	276	18	AAAR30311	Human tissue facto
31	194.5	6.4	276	19	AAAR61535	Human tissue facto
32	194.5	6.4	304	10	AAAP92002	Lipoprotein-associ
33	194.5	6.4	304	16	AAAR1884	Tissue factor path
34	194.5	6.4	304	16	AAAR67994	Human lipoprotein
35	194.5	6.4	304	20	AAAY49557	Human tissue facto
36	194.5	6.4	304	21	AAAY70273	Human tissue facto
37	194.5	6.4	352	17	AAAR92011	UBiquitin-TFPI fus
38	192.5	6.4	277	14	AAAR37312	Non-glycosylated T
39	192	6.4	170	18	AAAR30041	Human placental bl
40	192	6.4	170	21	AAAB14189	Human placental bl
41	192	6.4	170	21	AAAB14190	Human placental bl
42	192	6.4	179	18	AAAR30053	Human placental bl
43	192	6.4	179	21	AAAB14159	Human placental bl
44	192	6.4	197	18	AAAR30043	Mature human place
45	192	6.4	197	21	AAAB14160	Human placental bl

ALIGNMENTS

RESULT	1
AAAB8333	standard; Protein: 216 AA.
ID	AAAB8333
XX	AAAB8333:
XX	23-MAY-2001 (first entry)
DT	Human membrane or secretory protein clone PSEC0040.
DE	Human, secretory protein; membrane protein; vaccine; gene therapy;
XX	Rheumacoid arthritis; diabetes.
KW	Homo sapiens.
OS	EP1067182-A2.
XX	10-JAN-2001.
PD	07-JUL-2000; 2000EP-0114090.
XX	08-JUL-1999; 99UP-0194179.
PR	11-JAN-2000; 2000JP-0118775.
PR	02-MAY-2000; 2000JP-0183766.
XX	(HELI-) HELIX RES INST.
PA	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI	WPI: 2001-093989/11.
DR	N-PSDB: AAF93760.
XX	Nucleic acids encoding secretory proteins/membrane proteins, useful in
DR	gene therapy or as candidate target molecules in drug development -
PT	

XX Claim 1; SEQ ID 34; 609pp + CD ROM; English.
 PS This invention relates to nucleic acid sequences AAF93744 - AAF93916
 XX which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 XX Sequence 216 AA;
 SQ

Query Match 27.9%; Score 840; DB 22; Length 216;
 Best Local Similarity 71.5%; Pred. No. 7e-47;
 Matches 166; Conservative 4; Mismatches 29; Indels 34; Gaps 7;

QY 1 MPALRPLPLLLRLTSGALPLGSHPGVCPNOSPMLMVAAGTCEKCRDDCA 60
 Db 1 mpaLrPlpLlLlRlTsgAlPlGshPgVcPnOsPmLmVAGtCEKCRDDCA 60
 QY 61 AAEECCINVGCLHSCVAAARPPGSAAPPTAASCEGFVCPDQSGDDTDWQAPVCRKDRDC 120
 Db 61 aaEECCINvgCLhScVaaARPPGSAAPPTAAscEGfVcPDqSGDDTDWqAPVCRKDRDC 120
 QY 121 EKPSFSCASDGLTYNRCVMDAECRLGHLHIVPCRHVLSMPSP--SGPREPTARRPT 178
 Db 121 eKpSfScASdGLtYnRCvMDaECrLGhLhIVpCRhVLSMPSP--SGPREPTARRPT 178
 QY 179 PGAAPVPALYSSPPOAVGVGTASLHCDVSGRPPAVTWKOSHREMLINRP 233
 Db 173 pLpAp-----hLgPrP-----cLlPrCta-----hrrCrLgYrr 203

RESULT 2
 AAY84381
 ID AAY84381 standard; Protein; 59 AA.
 AC AAY84381;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE Amino acid sequence o a serine proteinase inhibitor designated ZKUN6.
 XX

KW Serine proteinase inhibitor; kunitz domain; ZKUN6; acute pancreatitis;
 KW cardiopulmonary bypass-induced pulmonary injury; myocardial infarction;
 KW allergy-induced protease release; deep vein thrombosis; shock;
 KW hyperfibrinolytic haemorrhage; emphysema; rheumatoid arthritis;
 KW adult respiratory distress syndrome; chronic inflammatory bowel disease;
 KW psoriasis; inflammatory conditions; platelet function; fibrinolysis;
 KW organ preservation; wound healing; haemostasis imbalance; cirrhosis;
 KW acquired coagulopathy; primary fibrinolysis; thrombolytic therapy;
 KW blood coagulation; proteolytic tissue degradation; gene therapy.
 KW
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH Disulfide-bond 6..56
 FT Domain 14...20
 FT Disulfide-bond 15...39
 FT Disulfide-bond 31...52
 PN WO200014235-A1.
 XX 16-MAR-2000.
 PD
 XX
 XX 01-SEP-1999; 99MO-US20202.
 PR 03-SEP-1998; 98US-0148092.
 XX (ZYMO) ZYMOGENETICS INC.
 PA ConKlin DC;
 XX
 PI
 XX WPI; 2000-256985/22.
 DR N-PsDB; AAZ9802, AAZ99803.
 XX
 PT Novel proteinase inhibitor Zkun6 comprising a kunitz domain, useful in
 PT the treatment or prevention of conditions associated with excessive
 PT proteinase activity
 XX

PS Claim 4; Page 40; 48pp; English.

CC The present sequence represents a serine proteinase inhibitor containing
 CC a kunitz domain, designated ZKUN6. The Zkun6 polypeptides are used in
 CC the treatment and prevention of conditions associated with excessive
 CC proteinase activity. The conditions include acute pancreatitis,
 CC cardiopulmonary bypass-induced pulmonary injury, allergy-induced
 CC protease release, deep vein thrombosis, myocardial infarction,
 CC shock, hyperfibrinolytic haemorrhage, emphysema, rheumatoid arthritis,
 CC adult respiratory distress syndrome, chronic inflammatory bowel disease,
 CC psoriasis, inflammatory conditions, platelet function, organ
 CC preservation, and wound healing. They are also useful in the treatment
 CC of conditions arising from haemostasis imbalance, including acquired
 CC coagulopathies, primary fibrinolysis and fibrinolysis due to cirrhosis,
 CC and complications of high-dose thrombolytic therapy. Zkun6 polypeptides
 CC may also be used for inhibiting blood coagulation in mammals, and for
 CC blockade of proteolytic tissue degradation. The Zkun6 polynucleotides
 CC are used in gene therapy to treat the above diseases. Transgenic animals,
 CC engineered to express Zkun6, and knockout animals with an absence of
 CC Zkun6 function, are used to study the Zkun6 gene and the encoded
 CC protein. They are useful for investigating the role of Zkun6 polypeptides
 CC in early development.
 CC
 XX Sequence 59 AA;
 SQ

Query Match 11.9%; Score 358; DB 21; Length 59;
 Best Local Similarity 100.0%; Pred. No. 1.9e-16;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 GFGDACVLPAYVGGPCRGWEPRMAYSPLLQOCHPEFYVGGCEGNGNMFHRESCEDACPVP 412
 Db 1 gPgDaCvLPaYvGGpCRGwEPRMAYSPllLQOChPEfYvGGCEgNGnMFhREsCEDaCPvP 59

RESULT 3
 AAR62523
 ID AAR62523 standard; peptide; 560 AA.
 XX
 AC AAR62523;
 XX

DT 06-JUN-1995 (first entry)
 XX

KW Hookworm anticoagulant.
 DE Hookworm; anticoagulant; serine protease-inhibitor; blood loss;
 XX

ID	AA092237	standard; protein; 147 AA.
AC	AA092237;	
DT	27-SEP-1996	(first entry)
DE	Human wild-type urinary trypsin inhibitor.	
KW	UTR; Kunitz domain; urinary trypsin inhibitor; elastase inhibitor;	
KW	recombinant protein production; yeast host cell; Pichia;	
KW	site-directed mutagenesis.	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	peptide	1..21
FT		/label= signal_peptide
FT	Domain	26..76
FT		/label= kunitz_domain_1
FT	Misc-difference	36..40
FT		/label= P1-P4'
FT		/note= "motif replaced by Ile-Ala-Phe-Phe-Pro to improve elastase inhibitor activity"
FT	Domain	78..145
FT		/label= kunitz_domain_2
FT	Disulfide-bond	26..76
FT	Disulfide-bond	35..59
FT	Disulfide-bond	51..72
FT	Disulfide-bond	82..132
FT	Disulfide-bond	91..115
FT	Disulfide-bond	107..128
FT	Modified-site	10
FT		/label= O-linked_glycosylation
FT	Modified-site	45
FT		/label= N-linked_glycosylation
PN	WO9603503-A1.	
PD	08-FEB-1996.	
PF	21-JUL-1995;	95WO-JP01449.
PR	21-JUL-1994;	94JP-0169221.
PA	(GREC) GREEN CROSS CORP.	
PI	Goto T, Horii H, Ideno S;	
PI	WPI: 1996-117048/12.	
PT	Production of recombinant urinary trypsin inhibitor in Pichia sp.	
PT	also Kunitz domain fragments of the inhibitor and new variants	
PT	having improved elastase inhibitor activity	
PS	Disclosure; Fig 2; 97pp; Japanese.	
CC	Recombinant urinary trypsin inhibitor (rUTI) can be produced in	
CC	Pichia yeast. By mutating a 5 amino acid motif (P1-P4') within kunitz	
CC	domain 1 (from MGMS to IAFPP), the resulting rUTI has improved	
CC	elastase inhibitory activity. The present sequence is that of	
CC	wild-type precursor UTI.	
Sequence	147 AA;	

QY	366	GPCRMEPRMAYSPILQGCHEFYVGGCGEGRNNFHSRRESCEDACPP	412
Db	89	gpcrafiglwaifaavkqkcvllfpygcgqngnkfysekecreycv	135
RESULT	6		
AAW25928			
ID	AAW25928	standard; Protein; 147 AA.	
XX	AC		
XX	AAW25928;		
XX			
DT	11-NOV-1997	(first entry)	
XX			
DE	Anti-trypsin inhibitor UTI.		
XX			
KW	Trypsin inhibitor; Kunitz domain; protease; active site; elastase;		
KW	neutrophil; disease; modification; site directed mutagenesis.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Modified-site	10	
FT	Domain	/label= O-linked glycosylation site	
FT		22..77	
FT	Disulfide-bond	/note= "Kunitz domain 1"	
FT	Disulfide-bond	26..76	
FT	Disulfide-bond	35..59	
FT	Active-site	36	
FT		/note= "active site residue"	
FT	Disulfide-bond	51..72	
FT	Domain	78..147	
FT		/note= "Kunitz domain 2"	
FT	Disulfide-bond	82..132	
FT	Disulfide-bond	91..115	
FT	Active-site	92	
FT		/note= "active site residue"	
XX	Disulfide-bond	107..128	
PN	JP09124700-A.		
XX			
XX	13-MAY-1997.		
XX			
PF	07-NOV-1995;	95JP-0288527.	
XX			
PR	07-NOV-1995;	95JP-0288527.	
XX			
PA	(GREC) GREEN CROSS CORP.		
XX			
DR	WPI; 1997-316576/29.		
XX			
DR	N-PSDB; AAT78950.		
XX			
PT	New protease inhibitor - useful for treating diseases involving		
PT	elastase		
XX			
PS	Disclosure; Fig 3; 37pp; Japanese.		
XX			
CC	This is the amino acid sequence of the anti-trypsin inhibitor UTI, which		
CC	is a protein containing 2 Kunitz domains. The nucleotide sequence		
CC	encoding the protein was used to construct a novel protease inhibitor by		
CC	replacing the active site of the Kunitz domain 1 (see AAT79080, AAT79081		
CC	and AAT79083). The modified protease inhibitors are targeted to the		
CC	disease elastase, especially from neutrophils and can be used to treat		
CC	protease associated with elastase. Modifications of the active site were		
CC	done by site directed mutagenesis.		
XX			
Sequence	147 AA;		
XX			

	Query Match	7.1%	Score 213;	DB 17;	Length 147;	
	Best Local Similarity	40.2%;	Pred No.	11e-06;		
	Matches	43; Conservative	12;	Mismatches	46;	Indels 6; Gaps 2
QY	CTPTSPHYLYMHYDQRGCGMTFPRARCDDGARKEFYETECGGACGACRGSDACVLPANQ	365				
Dh	35 cmgmstaryf----yngtsmactctfygscmgsmgnmlfvekecklcttcta--acnlpivyr	88				

Query Match	7.1%;	Score 213;	DB 18;	Length 147;
Best Local Similarity	40.2%;	Pred No. 1.1e-06;		
Matches	43;	Conservative	12;	Mismatches 46;
			Indels	6;
			Gaps	2

```

Oy 306 CTGPTSPHLVLMHYDQRGCGMTFPARGCDGAARGFETYEACQACARGDACLPAVO 365
    | | | | | : | | | | | : | | | | | : | | | | | :
Db 35 cmgmstsrYf-----yngtsmacetfgygcmgngnfvtekeclqctcrta--acnlpivY 88
Oy 366 GPCRGMPEPRMAYSPLLQOCHPEYVGGCEGNGNHNHRESCEDACPVP 412
    | | | | | : | | | | | : | | | | | : | | | | | :
Db 89 gpcrafiqlwatdavrkykcvlfpyggcgngnkfysekecreycgvp 135

RESULT 7
AAPB1110
ID AAPB1110 standard; protein: 352 AA.
XX
AC AAPB1110;
XX
DT 06-DEC-1990 (first entry)
XX
DE Sequence of new fusion protein contg. alpha-1-microglobulin (AMG)
DE and the HI-30 region of inter-alpha-trypsin inhibitor (III) light chain.
XX
KW Serine protease; enzyme; pancreatitis; atherosclerosis;
KW chronic inflammation; therapy; elastase.
XX
OS Homo sapiens.
XX
FH Key location/qualifiers
FT Protein 20..202
FT /label=AMG
FT Protein 206..350
FT /label=HI-30
FT Domain 226..282
FT /label=I
FT Domain 283..352
FT /label=II
FT Misc-difference 291..292
FT /note="Differs from the protein sequence of HI-30
FT purified from urine"
FT Misc-difference 343
FT /note="Differs from the protein sequence of HI-30
FT purified from urine"
XX
PN EP255011-A.
XX
PD 03-FEB-1988.
XX
PE 20-JUL-1987; 87EP-0110461.
XX
PR 29-JUL-1986; 86US-0891469.
XX
PA (MILE ) MILES LABORATORIES INC.
XX
PI Kaumeyer JF, Kotlick MP, Polazzi JO;
XX
DR WPI; 1988-030262/05.
XX
DR N-PSDB; AAN81432.
XX
PT New DNA sequence coding for fusion protein contg. alpha-microglobulin -
PT and inter-alpha-trypsin inhibitor, useful for treating excessive
PT elastase prodn.
XX
PS Disclousre; ; P; English.
XX
CC A fusion protein of the IRI light chain comprising AMG and HI-30 is
CC claimed. IRI is serine protease, potentially used for treating excessive
CC release of hydrolytic enzymes, esp. elastase, in conditions such as
CC pancreatitis, atherosclerosis and chronic inflammation.
XX
SQ Sequence 352 AA;

```

```

Oy 306 CTGPTSPHLVLMHYDQRGCGMTFPARGCDGAARGFETYEACQACARGDACLPAVO 365
    | | | | | : | | | | | : | | | | | : | | | | | :
Db 240 cmgmstsrYf-----yngtsmacetfgygcmgngnfvtekeclqctcrta--acnlpivY 293
Oy 366 GPCRGMPEPRMAYSPLLQOCHPEYVGGCEGNGNHNHRESCEDACPVP 412
    | | | | | : | | | | | : | | | | | : | | | | | :
Db 294 gpcrafiqlwatdavrkykcvlfpyggcgngnkfysekecreycgvp 340

RESULT 8
AAU14343
ID AAU14343 standard; Protein: 352 AA.
XX
AC AAU14343;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #214.
XX
KW Human; novel protein; Antianemic; osteopathic; antiinflammatory;
KW immunomodulatory; cyostatic; neuroprotective; vulnerrary; nocitropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; Immune disorder.
XX
KW Homo sapiens.
XX
OS WO200155437-A2.
XX
PN 02-AUG-2001.
XX
PD 25-JAN-2001; 2001WO-0502623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HSE-) HSEQ INC.
XX
PI Tang YF, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
XX
DR N-PSDB; AAS22648.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Example 4; Page 642-643; 894pp: English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicite an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,

```

Query Match 7.1%; Score 213; DB 9; Length 352;
 Best Local Similarity 40.2%; Pred. No. 3e-06;
 Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
the proliferation of cells.

Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, stroke, immune deficiencies resulting from bacterial, viral or

CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis.

CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.

XX
SQ Sequence 352 AA;

Query Match	7.18;	Score 213;	DB 22;	Length 352;
Best Local Similarity	40.28;	Pred. No. 3e-06;		

Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2

306 CUGPTSPHLVLMHYDDPQAGGCMTEPARGCDGARGFETYEACQACARPGDACVLEPAVQ 365

240 cmgmtsryf----yngtsmacetfgygcmgungnfvtckclqfcttva--acnlpivr 293

366 GPCRGMEPRNAVSLPQQCHPFVYGGCEGNGNNFHSRSCDACPVP 412

294 gpcratlqlwatdavkqcvlfpY9gcqgnkfysekecreycgvp 340

```
RESULT 10
0A005005
```

AAE05095 standard; Protein, 352 AA.

AAE05095;

18-SEP-2001 (first entry)

Human inter-alpha trypsin inhibitor (ITI) light chain.

Diagnosis; tumour; cancer; central nervous system; CNS; mab 69.31; monoclonal antibody; metastasis

monoclonal antibody; astrocytoma; serine protease inhibitor; primary brain tumour; brain metastasis; leukaemia; carcinoma; glioma; human interleukin-6 transmembrane inhibitor

glycine, alpha chain; ITI;
plasmin inhibitor.

Homo sapiens.

Key	Location/Qualifiers
Peptide	1 19

Protein	Signal_peptide
20	352

Active-site	/label= Mature_ITI_light_chain
20..32	

Binding-site	215	Forms part of the active site"
--------------	-----	--------------------------------

note="this residue is involved in binding to ITI
heavy chain via chondroitin sulfate"

Active-site	241..242	chain via chromodomain sulphate
		note="Forms part of the active site"

Active-site	297..298	Forms part of the active site" /note=
Active-site	297..298	Forms part of the active site" /note=

WO200153835-A2.

26-JUL-2001. X

24-JAN-2001; 2001WO-US02269.

24-JAN-2000; 2000US-0491479.

(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

Lim Y, Hixson DC;
X

08-MAR-2000; 2000WO-US05882.

antitubercular; antidiarrhoeal; antihelminthic; antitumour; antitubercular; antidiarrhoeal; antihelminthic; antiviral; antinflammatory; antihypertoid; antiallergic; antibacterial; dermatological; neuroprotective; cardiant; thrombolytic; coagulant; neurologic; vasotropic; antiparasitic and antiparasitic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.

KM Drug; cell membrane-directed drug; phospholipid; lipid bilayer;
KW cell cortex; blood coagulation; inflammation; immunological disorder

XX
FN
XX

05-JAN-1998; 98WO-JP000002.


```

XX      Trypsin inhibitor: Kunitz domain; protease; active site; elastase;
KW      neutrophil; disease; modification; site directed mutagenesis.
XX      Synthetic.
OS
XX      Key
FH      Location/Qualifiers
FT      1..20
FT      Peptide
FT      /note= "signal peptide"
FT      Protein
FT      21..144
FT      /note= "mature protein"
FT      Domain
FT      21..76
FT      /note= "Kunitz domain 1"
FT      Disulfide-bond
FT      25..75
FT      Disulfide-bond
FT      34..58
FT      Misc-difference
FT      35..39
FT      /note= "mutated region; amino acids substituted for
FT      wild type amino acid sequence MGMTS; this
FT      includes the active site residue of domain 1
FT      (aa 36 of the wild type sequence AAW25928)"
FT      Disulfide-bond
FT      50..71
FT      Misc-difference
FT      51
FT      /note= "mutated amino acid; replaces amino acid Glu in
FT      wild type sequence"
FT      Misc-difference
FT      54
FT      /note= "mutated amino acid; replaces amino acid Gln in
FT      wild type sequence"
FT      Domain
FT      77..144
FT      /note= "Kunitz domain 2"
FT      Disulfide-bond
FT      81..131
FT      Disulfide-bond
FT      90..114
FT      Active-site
FT      91
FT      /note= "active site residue"
FT      Disulfide-bond
FT      106..127
XX      JP09124700-A.
XX      PN
XX      PD
XX      13-MAY-1997.
XX      PF
XX      07-NOV-1995; 95JP-0288527.
XX      PR
XX      07-NOV-1995; 95JP-0288527.
XX      PA
XX      (GREC ) GREEN CROSS CORP.
XX      DR
XX      WPI: 1997-316576/29.
XX      DR
XX      N-PSDB; AAT79086.
XX      PT
XX      New protease inhibitor - useful for treating diseases involving
XX      elastase
XX      PS
XX      Disclosure; Fig 30; 37pp; Japanese.
XX      CC
XX      This is the amino acid sequence of the polypeptide encoded by the insert
XX      in plasmid pHR337 which comprises the novel elastase specific inhibitor
XX      Epi-d21-RPDF-52-55 (AAW25932). The inhibitor sequence is linked
XX      downstream of the yeast invertase (SUC2) signal peptide sequence. The
XX      modified protease inhibitors are targeted to the protease elastase,
XX      especially from neutrophils and can be used to treat diseases associated
XX      with elastase. Modifications of the active site were done by site
XX      directed mutagenesis.
XX      CC
XX      Sequence 144 AA;
SQ

```

```

Query Match          7.0%; Score 210; DB 18; Length 144;
Best Local Similarity 39.0%; Pred. No. 1.7e-06;
Matches 41; Conservative 12; Mismatches 50; Indels 2; Gaps 1;

```

```

OY      308 GPTSPHLVLMHYDPQRCGCMFPFARCGDCAARGFETYEACCOACAGPDAVLPVQGP 367
        || : || : || : || : || : || : || : || : || : || : || : || : ||
DB      32 gpciaffpyfyngtsmaeqftfyggcmgnfnfvekecdlctcrva--acnlpivrgp 89

```

```

OY      368 CRGMEPRWAYSPLLQCHPEVYGCCEGNGNPNHRSRESCEDACVP 412
        || : || : || : || : || : || : || : || : || : || : || : || : ||
DB      90 crafqlwaifavkqkcvlfpfygqcgngnkfysekecreycgvp 134

RESULT 15
AAW25935
ID      AAW25935 standard; Protein; 145 AA.
XX
XX      AAW25935;
XX
XX      11-NOV-1997 (first entry)
XX
XX      Novel protease inhibitor Epi-UTR-RPDF-52-55.
XX
XX      Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW      neutrophil; disease; modification; site directed mutagenesis.
XX      Synthetic.
OS
XX      Key
FH      Location/Qualifiers
FT      Modified-site
FT      10
FT      /label= "O-linked glycosylation site"
FT      Domain
FT      22..77
FT      /note= "Kunitz domain 1"
FT      Misc-difference
FT      22..25
FT      /note= "mutated amino acids; replaces wild type
FT      sequence KEDS"
FT      Disulfide-bond
FT      26..76
FT      Disulfide-bond
FT      35..59
FT      Misc-difference
FT      15..19
FT      /note= "mutated region; amino acids substituted for
FT      wild type amino acid sequence MGMTS; this
FT      includes the active site residue of domain 1"
FT      Disulfide-bond
FT      51..72
FT      Misc-difference
FT      52
FT      /note= "mutated amino acid; replaces amino acid Glu in
FT      wild type sequence"
FT      Misc-difference
FT      55
FT      /note= "mutated amino acid; replaces amino acid Gln in
FT      wild type sequence"
FT      Domain
FT      78..147
FT      /note= "Kunitz domain 2"
FT      Disulfide-bond
FT      82..132
FT      Disulfide-bond
FT      91..115
FT      Active-site
FT      92
FT      /note= "active site residue"
FT      Disulfide-bond
FT      107..128
XX      JP09124700-A.
XX      PN
XX      PD
XX      13-MAY-1997.
XX      PF
XX      07-NOV-1995; 95JP-0288527.
XX      PR
XX      07-NOV-1995; 95JP-0288527.
XX      PA
XX      (GREC ) GREEN CROSS CORP.
XX      DR
XX      WPI: 1997-316576/29.
XX      DR
XX      N-PSDB; AAT79083.
XX      PT
XX      New protease inhibitor - useful for treating diseases involving
XX      elastase
XX      PS
XX      Claim 8; Page 4; 37pp; Japanese.
XX      CC
XX      This is the amino acid sequence of a protease inhibitor construct based
XX      on the anti-trypsin inhibitor UTR (AAW25928). The novel inhibitor
XX      contains the Kunitz domains 1 and 2 of UTR with several amino acid
XX      substitutions: amino acids 22-25 of the wild type protein (KEDS) are
XX      replaced by the sequence RPDF; the active site sequence and adjacent
XX      amino acids (amino acids 36-40 of the wild type sequence) are replaced by

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:51:44 ; Search time 45.36 seconds
(without alignments)
10.076 Million cell updates/sec

Title: US-09-819-136-2_COPY_525_530

Perfect score: 6

Sequence: 1 EKRVRK 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	100.0	166	2 T29929	hypothetical prote
2	6	100.0	230	2 C56657	PFEMP2/MESA (clone
3	6	100.0	309	2 T43752	site-specific DNA
4	6	100.0	329	2 G81344	ABC transport syst
5	6	100.0	691	2 G69724	DNA topoisomerase
6	6	100.0	984	2 G86393	protein T24P13.6
7	6	100.0	1526	2 A45605	mature-parasite-in
8	6	100.0	28	2 T38041	similarity to yeast
9	6	100.0	52	2 S43798	hypothetical prote
10	6	100.0	67	2 I53485	GRP-binding protei
11	6	100.0	74	2 D69353	hypothetical prote
12	6	100.0	80	2 D69492	conserved hypotet
13	6	100.0	95	2 T43006	HMG protein 1.1 -
14	6	100.0	99	2 C84022	hypothetical prote
15	6	100.0	119	2 C49923	ribosomal protein
16	6	100.0	126	2 A56657	PFEMP2/MESA protei
17	6	100.0	129	2 T64094	ribosomal protein
18	6	100.0	131	2 B84988	30S ribosomal prot
19	6	100.0	144	2 D84786	hypothetical prote
20	6	100.0	144	2 G69255	hypothetical prote
21	6	100.0	145	2 H69322	transcription regu
22	6	100.0	147	2 T09722	histone H2B1 - upl
23	6	100.0	147	2 S73407	hypothetical prote
24	6	100.0	151	2 A39392	Rab6 DNA-repair ho
25	6	100.0	151	2 D84688	probable histone H
26	6	100.0	154	2 I47153	transcription fact
27	6	100.0	156	2 C64390	probable transcript
28	6	100.0	162	2 T26100	hypothetical prote
29	6	100.0	164	2 B72364	funarate hydratase

30	5	83.3	165	2 C69065	hypothetical prote
31	5	83.3	167	2 T04903	iron-sulfur cofact
32	5	83.3	167	2 G70436	hypothetical prote
33	5	83.3	171	2 A36958	DnaJ homolog (hscA
34	5	83.3	171	2 E85897	hypothetical prote
35	5	83.3	181	2 S53855	ribosomal protein
36	5	83.3	182	2 S29611	pollen-preferentia
37	5	83.3	189	2 H70323	hypothetical prote
38	5	83.3	204	2 T45948	hypothetical prote
39	5	83.3	205	2 E49316	transcription anti
40	5	83.3	206	2 F71450	hypothetical prote
41	5	83.3	210	2 B75450	hypothetical prote
42	5	83.3	218	2 A43958	GRP-binding protei
43	5	83.3	218	2 G66307	probable glutathio
44	5	83.3	219	2 B29224	GRP-binding protei
45	5	83.3	219	2 D34323	GRP-binding protei

ALIGNMENTS

RESULT 1

T29929

hypothetical protein F2989.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T29929

R:Goela, D.; Gattung, S.

submitted to the EMBL Data Library, September 1996

A:Description: The sequence of C. elegans cosmid F29B9.

A:Reference number: 220710

A:Accession: T29929

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-166 <GO>

A:Cross-references: EMBL:U70849; PIDN:AA09118.1; GSPDB:GN00022; CESP:F29B9.6

A:Experimental source: strain Bristol N2; clone F29B9

C:Genetics:

A:Gene: CESP:F29B9.6

A:Map position: 4

A:Introns: 18/3; 59/3; 90/1; 141/2

C:Superfamily: human ubiquitin--protein ligase E2

Query Match 100.0% Score 6; DB 2; Length 166;

Best Local Similarity 100.0%; Pred. No. 6.8;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKRVRK 6
DB 145 EKRVRK 150

RESULT 2

C56657

PFEMP2/MESA (clone 9025/60) - Plasmodium falciparum (fragment)

C:Species: Plasmodium falciparum

C:Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 03-Nov-2000

C:Accession: C56657

R:Saul, A.; Yeganeh, F.; Howard, R.J.

Immunol. Cell Biol. 70, 353-355, 1992

A>Title: Conservation of repeating structures in the PFEMP2/MESA protein of Plasmodiu

A:Reference number: A56657; MUID:93122844

A:Contents: Malayan Camp isolate

A:Accession: C56657

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <SAU>

A:Cross-references: GB:S52459; NID:9263279; PID:9263280

A>Note: sequence extracted from NCBI backbone (NCBIN:122452, NCBI:122455)

C:Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homolog

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 230;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
DB 59 EKRVRK 64

NO

RESULT 3
T43752
site specific DNA endonuclease homolog [imported] - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000
C:Accession: T43752
R:Ogawa, S.; Yoshino, R.; Angata, K.; Pi, M.; Iwamoto, M.; Kuroe, K.; Matsuo, K.; Morio, submitted to the EMBL Data Library, December 1996
A:Description: The mitochondrial DNA of Dictyostelium discoideum. Complete sequence, gen
A:Reference number: 222666
A:Accession: T43752
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-309 <OGA>
A:Cross-references: EMBL:AB000109; PIDN:BAAT8056.1
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: COI intron 9 protein homology
C:Keywords: mitochondrion
F:25-251/Domain: COI intron 9 protein homology <CI9>

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 309;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
DB 275 EKRVRK 280

NO

RESULT 4
G81344
ABC transport system ATP-binding protein Cj0732 [imported] - Campylobacter jejuni (strain
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: G81344
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barre Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf
A:Reference number: A81250; MUID:20150912
A:Accession: G81344
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <PAR>
A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB73006.1; PID:g696818
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0732

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 329;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
DB 111 EKRVRK 116

NO

RESULT 5
G69724
DNA topoisomerase I topA - Bacillus subtilis

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G69724

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Acevedo, V.; Ber A.; Ehrlich, S.D.; Emerson, P.T.; Enliam, K.D.; Errington, J.; Fabelt, C.; Ferrari, Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gal lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele Rieger, M.; Rivolta, C.; Kocha, E.; Koche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani A:Authors: Schlecht, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se akewuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstine, P.; Tognoni, A.; Tosato, V.; Uchiya T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: G69724
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-691 <KUN>
A:Cross-references: GB:299112; GB:AL009126; NID:g2633902; PIDN:CAB13485.1; PID:g26339
A:Experimental source: strain 168
C:Genetics:
A:Gene: topA
C:Superfamily: DNA topoisomerase I

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 691;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
DB 555 EKRVRK 560

NO

RESULT 6
G86393
Protein T24P13.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86393
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.E.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marzla Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G86393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-984 <STO>
A:Cross-references: GB:AE005172; NID:g9295721; PIDN:AA87027.1; GSPDB:GN00141
C:Genetics:
A:Gene: T24P13.6
A:Map position: 1

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 984;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
DB 597 EKRVRK 602

RESULT 7
AA5605
mature-parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: AA5605; A54517
R:Coppel, R. L.
Mol. Biochem. Parasitol. 50, 335-347, 1992
A:Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human erythrocytes
A:Reference number: AA5605; MUID:92158014
A:Accession: AA5605
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <COP>
A:Experimental source: Papua New Guinea isolate FC27
A:Note: sequence extracted from NCBI backbone (NCBIN:83648, NCBI:83656)
R:Coppel, R. L.; Cuivemor, J. G.; Bianco, A. E.; Crewther, P. E.; Stahl, H. D.; Brown, G. V.; Mol. Biochem. Parasitol. 20, 265-277, 1986
A:Title: Variable antigen associated with the surface of erythrocytes infected with mature-parasite-infected erythrocyte surface antigen; tandem repeat
A:Reference number: A54517; MUID:87014571
A:Accession: A54517
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 222-443, 'K', 445 <COP>
A:Cross-references: GB:M15319; NID:9160060; PID:9552170
C:Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology
C:Keywords: surface antigen; tandem repeat

Query Match 100.0%; Score 6; DB 2; Length 1526;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
DB 578 EKRVRK 583

RESULT 8
T38041
similarity to yeast mating protein SSF1 - fission yeast (Schizosaccharomyces pombe) (first)
C:Species: Schizosaccharomyces pombe
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T38041
R:McDougal, R. C.; Rajandream, M. A.; Barrett, B. G.; Whitehead, S.; Churcher, C. M. submitted to the EMBL Data Library, August 1999
A:Reference number: 221764
A:Accession: T38041
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-28 <MCD>
A:Cross-references: EMBL:AL109951; PIDN:CAB53054.1; GSPDB:GN00066; SPDB:SPAC1B9.01c
A:Experimental source: strain 972h-; cosmid c1B9
C:Genetics:
A:Gene: SPDB:SPAC1B9.01c
A:Map position: 1

Query Match 83.3%; Score 5; DB 2; Length 28;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6
DB 7 KRVKK 11

RESULT 9
S43798
hypothetical protein 34 - phage SPPI
C:Species: phage SPPI

C:Date: 25-Dec-1994 #sequence_revision 26-May-1995 #text_change 11-May-2000
C:Accession: S43798; T42330; S41172
R:Pedre, X.; Weise, F.; Chal, S.; Lueder, G.; Alonso, J. C.
J. Mol. Biol. 236, 1324-1340, 1994
A:Title: Analysis of cis and trans acting elements required for the initiation of DNA replication in phage SPPI
A:Reference number: S43798; MUID:94172631
A:Accession: S43798
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-52 <PED>
A:Cross-references: EMBL:X67865; NID:9472886; PIDN:CAA48049.1; PID:9439629
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1992
R:Alonso, J. C.; Lueder, G.; Stege, A. C.; Chal, S.; Weise, F.; Trautner, T. A. Gene 204, 201-212, 1997
A:Title: The complete nucleotide sequence and functional organization of Bacillus subtilis phage SPPI
A:Reference number: 22137; MUID:98094274
A:Accession: T42330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-52 <ALO>
A:Cross-references: EMBL:X97918; PIDN:CAA6537.1

Query Match 83.3%; Score 5; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6
DB 5 KRVKK 9

RESULT 10
I53485
GTP-binding protein rab3B - rat (fragment)
C:Species: Rattus sp. (rat)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 19-Jan-2001
C:Accession: I53485
R:Oberhauser, A. F.; Balan, V.; Fernandez-Badilla, C. L.; Fernandez, J. M. FEBS Lett. 339, 171-174, 1994
A:Title: RT-PCR cloning of Rab3 isoforms expressed in peritoneal mast cells.
A:Reference number: I53485; MUID:94148079
A:Accession: I53485
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-67 <RES>
A:Cross-references: GB:S68807; NID:9545388; PIDN:AAB29894.1; PID:9545389
C:Genetics:
A:Gene: rab3B
C:Superfamily: ras transforming protein; translation elongation factor Tu homology
C:Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop
F:14-21/Region: nucleotide-binding motif A (P-loop)

Query Match 83.3%; Score 5; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 5
DB 53 EKRVRK 57

RESULT 11
D69353
hypothetical protein AF0828 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: D69353
R:Klenk, H. P.; Clayton, R. A.; Tomb, J. F.; White, O.; Nelson, K. E.; Ketchum, K. A.; Dodson, J.; Fleischmann, R. D.; Quackenbush, J.; Lee, N. H.; Sutton, G. G.; Gill, S.; Kirsch, E. Glöck, A.; Zhou, L.; Overbeek, R.; Gokey, J. D.; Weidman, J. F.; McDonald, L. Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Moese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343
 A:Accession: D69353
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-74 <KLE>
 A:Cross-references: GB:AE001047; GB:AE000782; NID:g2689370; PIDN:AAB90423.1; PID:g264978

Query Match

Best Local Similarity 83.3%; Score 5; DB 2; Length 74;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVK 6
 DB 29 KRVK 33

RESULT 12

D69492

conserved hypothetical protein AF1941 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
 C:Accession: D69492

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsch, E.F.
 ; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 ; Nature 350, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Moese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-80 <KLE>
 A:Cross-references: GB:AE000969; GB:AE000782; NID:g2689292; PIDN:AAB89310.1; PID:g264859
 C:Superfamily: conserved hypothetical protein M01593

Query Match 83.3%; Score 5; DB 2; Length 80;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVK 6
 DB 31 KRVK 35

RESULT 13

T43006

HMG protein 1.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
 C:Accession: T43006

R:Kurtz, T.; Schulze, E.
 submitted to the EMBL Data Library, April 1998
 A:Description: The high mobility group proteins of Caenorhabditis elegans.
 A:Reference number: 222282
 A:Accession: T43006
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-95 <KUR>
 A:Cross-references: EMBL:AF056576; PIDN:AACT8598.1
 C:Genetics:

A:Gene: hmg-1.1

Query Match 83.3%; Score 5; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVK 6
 DB 14 KRVK 18

RESULT 14

C84022

hypothetical protein BH2979 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: C84022

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
 ; Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20263314
 A:Accession: C84022
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <STO>
 A:Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BA06698.1; GSPDB:G
 A:Experimental source: Strain C-125
 C:Genetics:
 A:Gene: BH2979

Query Match 83.3%; Score 5; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRK 5
 DB 95 EKRK 99

RESULT 15

C49923

ribosomal protein S13 - Chlamydia trachomatis
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
 C:Accession: C49923

R:Tan, M.; Klein, R.; Grant, R.; Ganem, D.; Engel, J.
 J. Bacteriol. 175, 7150-7159, 1993
 A:Title: Cloning and characterization of the RNA polymerase alpha-subunit operon of C
 A:Reference number: A49923; MUID:94042887
 A:Accession: C49923
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-119 <TAN>
 A:Cross-references: GB:L23478; NID:g385200; PIDN:AAA16205.1; PID:g385203
 A:Experimental source: mouse pneumonitis
 A:Note: sequence extracted from NCBI backbone (NCBIN:139888, NCBIF:139891)
 C:Superfamily: Escherichia coli ribosomal protein S13

Query Match 83.3%; Score 5; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRK 5
 DB 42 EKRK 46

Search completed: February 26, 2002, 01:51:46
 Job time: 455 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:53:16 ; Search time 78.18 seconds
(without alignments)
11.226 Million cell updates/sec

Title: US-09-819-136-2_COPY_525_530

Perfect score: 6

Sequence: 1 EKRVRKK 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	166	5	095017 caenorhabd1
2	6	100.0	230	5	006167 plasmodium
3	6	100.0	276	5	09vfe6 drosophila
4	6	100.0	309	8	021043 dictyostell
5	6	100.0	329	2	09pph5 campylobact
6	6	100.0	984	10	09LOY4 arabisidops
7	6	100.0	1510	5	025920 plasmodium
8	6	100.0	1661	5	006166 plasmodium
9	5	83.3	28	3	09vfe2 schizosacch
10	5	83.3	51	9	038591 bacterioph
11	5	83.3	52	9	038138 bacterioph
12	5	83.3	74	1	029430 archaeoglob
13	5	83.3	82	5	09GYZ6 trypanosoma
14	5	83.3	88	2	048711 lactococcus
15	5	83.3	89	11	P70506 rattus norv
16	5	83.3	95	5	077095 caenorhabd1
17	5	83.3	96	2	09K5E0 campylobact
18	5	83.3	98	2	09WMN9 nostoc sp.
19	5	83.3	99	2	09K8M7 bacillus ha

20	5	83.3	104	11	09D422 mus musculu
21	5	83.3	106	11	09D9C3 mus musculu
22	5	83.3	112	11	09GYT8 rattus norv
23	5	83.3	120	12	0907X1 human immun
24	5	83.3	122	12	064873 human adeno
25	5	83.3	126	5	006165 plasmodium
26	5	83.3	129	2	09CL52 pasteurella
27	5	83.3	139	10	09W3H6 cicer arlet
28	5	83.3	140	10	09YFE2 coryza sativ
29	5	83.3	142	5	09VDL0 drosophila
30	5	83.3	144	10	09SLU4 arabisidops
31	5	83.3	145	1	029671 archaeoglob
32	5	83.3	145	6	029012 us scrofa
33	5	83.3	151	10	09S196 arabisidops
34	5	83.3	153	5	09U7B1 macrotrache
35	5	83.3	164	2	09W7I1 thermotoga
36	5	83.3	165	1	027532 methanobact
37	5	83.3	165	5	09V535 drosophila
38	5	83.3	167	2	067521 aquifex aeo
39	5	83.3	167	10	049627 arabisidops
40	5	83.3	177	5	09N7I6 leishmania
41	5	83.3	179	4	09P069 homo sapien
42	5	83.3	182	10	043713 lilium long
43	5	83.3	189	2	066620 aquifex aeo
44	5	83.3	196	5	023205 caenorhabd1
45	5	83.3	198	5	077005 plasmodium

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	166 AA.
ID	095017			
AC	095017			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	UBIQUITIN-CONJUGATING ENZYME E2 (EC 6.3.2.19) (UBIQUITIN-PROTEIN			
DE	LIGASE) (UBIQUITIN CARRIER PROTEIN).			
GN	UBC-9.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;			
OC	Rhabditidae; Pelodierinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	I.A.T., Sun B., Lee M.-K., Teo T.-S.;			
RT	"ubc-9 of Caenorhabditis elegans: identification, characterization and			
RT	interaction with smt-3.";			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER			
CC	PROTEINS.			
CC	- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +			
CC	PYROPHOSPHATE + PROTEIN N-UBIQUITYLYSINE.			
CC	- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.			
CC	- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-			
CC	THIOLESTER FORMATION.			
CC	- SIMILARITY: TO UBIQUITIN-CONJUGATING ENZYMES DOMAIN.			
DR	EMBL: AF106565; AAC97374.1; -			
DR	HSSP: P50550: 1098.			
DR	InterPro: IPR000608; UBC-conjugat.			
DR	PIfam: PF00179; UQ_con: 1.			
DR	SMART: SM00212; UBC; 1.			
DR	PROSITE: PS00183; UBIQUITIN-CONJUGAT_1; UNKNOWN_1.			
DR	PROSITE: PS50127; UBIQUITIN-CONJUGAT_2; 1.			
KW	ligase; ubiquitin conjugation.			
FT	UBIQUITIN (BY SIMILARITY).			
SQ	SEQUENCE 166 AA; 19115 MW; E5A5AF73DE286C CRC64;			

Query Match 100.0%; Score 6; DB 5; Length 166;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
111111
Db 145 EKRVRK 150

NO

RESULT 2
ID 006167 PRELIMINARY; PRT; 230 AA.
AC 006167;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE ANTIMICROBIAL PROTEIN PFEPM2 (FRAGMENT).
OS Plasmodium falciparum (Isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93122844; PubMed=1478701;
RA Saul A., Yeganeh F., Howard R.J.;
RT "Conservation of repeating structures in the PTEM2/MESA protein of
RT Plasmodium falciparum."
RL Immunol. Cell Biol. 70:353-355(1992).
DR EMBL; S52459; AAB24870.1;
KW Antigen.
FT NON_TER
SQ SEQUENCE 230 AA; 26195 MW; 20582BF8AB542DCD CRC64;

Query Match 100.0%; Score 6; DB 5; Length 230;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
111111
Db 59 EKRVRK 64

NO

RESULT 3
ID 09VPE6 PRELIMINARY; PRT; 276 AA.
AC 09VPE6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE CG3817 PROTEIN.
GN CG3817.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolintsov S.,
RA Borkov D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honick J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzay D.M., Nelson D.L.,
RA Nelson D.R., Nelson K., Nixon K., Nusken D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003706; AAF55114.1;
DR FlyBase; FBgn0038275; CG3817.
SQ SEQUENCE 276 AA; 31206 MW; BC44FA61B769A0D6 CRC64;

Query Match 100.0%; Score 6; DB 5; Length 276;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6
111111
Db 6 EKRVRK 11

NO

RESULT 4
ID 021043 PRELIMINARY; PRT; 309 AA.
AC 021043;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE AT2A PROTEIN.
OS Dictyostelium discoideum (Slime mold).
OC Mitochondrion.
OC Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
OX NCBI_TaxID=44689;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-AX3;
RX MEDLINE=97354303; PubMed=9210597;
RA Ogawa S., Naito K., Angata K., Morio T., Urushihara H., Tanaka Y.,
RA Ogawa S., Matsuo K., Angata K., Yanagisawa K., Tanaka Y.;
RT "Group-I introns in the cytochrome c oxidase genes of Dictyostelium
RT discoideum: two related ORFs in one loop of a group-I intron, a coxI/2
RT hybrid gene and an unusually large cox3 gene."
RL Curr. Genet. 31:80-88(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AX3;
RX MEDLINE=97354303; PubMed=9210597;
RA Ogawa S., Naito K., Angata K., Morio T., Urushihara H., Tanaka Y.;
RT "A site-specific DNA endonuclease specified by one of two ORFs encoded
RT by a group I intron in Dictyostelium discoideum mitochondrial DNA."
RL Gene 191:115-121(1997).
DR EMBL; D50297; BAA21124.1;
DR EMBL; AB000109; BAA78056.1;
DR InterPro; IPR001982; Intron_endonuc.
DR Pfam; PF00961; Intron_maturase; 1.
KW Mitochondrion; Endonuclease.

SQ SEQUENCE 309 AA; 37116 MW; 2F7E4085CCF2FA30 CRC64;

Query Match 100.0%; Score 6; DB 8; Length 309;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6

DB 275 EKRVRK 280

RESULT 5

O9PH5

AC O9PH5; PRELIMINARY; PRT; 329 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE ABC_TRANSPORT_SYSTEM_ATP-BINDING_PROTEIN.

GN C30732.

OS Campylobacter jejuni.

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;

OC Campylobacter.

OX NCBI_TaxID=197;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCCTC 11168;

RA MEDLINE=20150912; PubMed=10688204;

RA Parthill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,

Baslam D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,

RA Jagals K., Chillingworth T., Moule S., Pallen M.J., Penn C.W.M.,

RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,

RA Whitehead S., Barrett B.G.;

RT "The genome sequence of the food-borne pathogen Campylobacter jejuni

reveals hypervariable sequences.";

RT Nature 403:665-668(2000).

CC -1 SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC

TRANSPORTERS).

CC EMBL: AL139076; CAB73006.1; -.

DR InterPro: IPR003593; AAA.

DR InterPro: IPR003439; ABC_transport.

DR InterPro: IPR001687; ATP_GTP_A.

DR Pfam: PF00005; ABC_tran; 1.

DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.

KW ATP-binding; Complete proteome; Transport.

SEQUENCE 329 AA; 37366 MW; 91EDC9759FB98408 CRC64;

Query Match 100.0%; Score 6; DB 2; Length 329;

Best Local Similarity 100.0%; Pred. No. 31;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6

DB 111 EKRVRK 116

RESULT 6

O9LOV4

AC O9LOV4; PRELIMINARY; PRT; 984 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE T24P13.6.

GN Arabidopsis thaliana (Mouse-ear cress).

OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Eurosid II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Johnson-Hopson C., Dunn P., Brooks S., Buehler E., Chao Q., Khan S.,

RA Kim C., Shinn P., Altafi H., Bel Q., Chin C., Chlou J., Choi E.,

RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,

RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N.,

RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Federspiel N.A.,

RA Theologis A., Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC T24P13 from chromosome

I.";

RT Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,

RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E.,

RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,

RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N.,

RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,

RA Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AC006535; AAF87027.1; -.

SEQUENCE 984 AA; 112142 MW; E8A9D2B864839BF CRC64;

Query Match 100.0%; Score 6; DB 10; Length 984;

Best Local Similarity 100.0%; Pred. No. 73;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVRK 6

DB 597 EKRVRK 602

RESULT 7

O25920

AC O25920; PRELIMINARY; PRT; 1510 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

GN MATURE-PARASITE-INFECTED ERYTHROCYTE SURFACE ANTIGEN.

OS MESA.

OC Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Hemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D10;

RA MEDLINE=92158014; PubMed=1741020;

RA Coppel R.;

RT "Repeat structures in a Plasmodium falciparum protein (MESA) that

binds human erythrocyte protein 4.1.";

RL Mol. Biochem. Parasitol. 50:335-347(1992).

DR EMBL: M69183; AA29651.1; -.

DR InterPro: IPR001623; DnaJ_N.

DR Pfam: PF00226; DnaJ_1.

DR PROSITE: PS50076; DnaJ_2; 1.

DR SMART: SM00271; DnaJ_1.

SEQUENCE 1510 AA; 177185 MW; 95C0F57FF76EDA9 CRC64;

Query Match 100.0%; Score 6; DB 5; Length 1510;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRYVK 6
 DB 562 KRYVK 567

RESULT 8
 006166

ID 006166 PRELIMINARY; PRT; 1661 AA.

AC 006166; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MATURE PARASITE-INFECTED ERYTHROCYTE SURFACE ANTIGEN (ANTIGENIC PROTEIN PREMP2).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PALO ALTO;
 RX MEDLINE=92158014; PubMed=1741020;
 RA Coppel R.L.;
 RT "Repeat structures in a plasmodium falciparum protein (MESA) that binds human erythrocyte protein 4.1.";
 RL Mol. Biochem. Parasitol. 50:335-347(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PALO ALTO;
 RX Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 797-850 FROM N.A.
 RX MEDLINE=93122844; PubMed=1478701;
 RA Saul A., Yeganeh F., Howard R.J.;
 RT "Conservation of repeating structures in the PfEMP2/MESA protein of Plasmodium falciparum.";
 RL Immunol. Cell Biol. 70:353-355(1992).
 DR EMBL; AF056936; AAC13303.1; -;
 DR EMBL; S52458; AAB24869.1; -;
 DR InterPro: IPR001623; DnaJ_N.
 DR Pfam: PF00226; DnaJ_1.
 DR PROSITE; PSS0076; DnaJ_2; 1.
 DR SMART; SM00271; DnaJ; 1.
 KW Antigen.
 SQ SEQUENCE 1661 AA; 195479 MW; AF340527D85A9D29 CRC64;

Query Match 100.0%; Score 6; DB 5; Length 1661;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRYVK 6
 DB 578 KRYVK 583

RESULT 9
 09UTP2 PRELIMINARY; PRT; 28 AA.

AC 09UTP2; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE SIMILARITY TO YEAST MATING PROTEIN SPS1 (FRAGMENT).
 GN SPACB12.01.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=972H-1;
 RA McDougall R.C., Rajendram M.A., Barrell B.G., Whitehead S.,
 RA Churcher C.M.;
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL109951; CAB53054.1; -;
 FT NON_TER 28
 SQ SEQUENCE 28 AA; 3059 MW; 465C7D7B2B3E0C27 CRC64;

Query Match 83.3%; Score 5; DB 3; Length 28;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRYVK 6
 DB 7 KRYVK 11

RESULT 10
 038591 PRELIMINARY; PRT; 51 AA.

AC 038591; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE PROMOTER 3.
 OS Bacteriophage SPPL.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85263531; PubMed=2991424;
 RA Taylor R., Bensl G., Morelli G., Canosi U., Trautner T.A.;
 RT "The genome of Bacillus subtilis phage SPPL: structure of an early promoter.";
 RL J. Gen. Microbiol. 131:1259-1262(1985).
 DR EMBL; M15348; AAA32603.1; -;
 SQ SEQUENCE 51 AA; 6251 MW; 9D13D86A73E7039E CRC64;

Query Match 83.3%; Score 5; DB 9; Length 51;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRYVK 6
 DB 5 KRYVK 9

RESULT 11
 038138 PRELIMINARY; PRT; 52 AA.

AC 038138; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE ORF34.
 OS Bacteriophage SPPL.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94172631; PubMed=8126723;
 RA Pedre X., Weise F., Chal S., Lueder G., Alonso J.C.;
 RT "Analysis of cis and trans acting elements required for the initiation of DNA replication in the Bacillus subtilis bacteriophage SPPL.";
 RL J. Mol. Biol. 236:1324-1340(1994).

RN [2]
 RP SEQUENCE FROM N.A.
 RA Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Alonso J.C.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X67865; CAA48049.1; -;
 DR EMBL: X97918; CAA66537.1; -;
 SQ SEQUENCE 52 AA; 6297 MW; 5BDD87222B47EFB1 CRC64;

Query Match 83.3%; Score 5; DB 9; Length 52;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVK 6
 |||||
 Db 5 KRVK 9

RESULT 12
 ID 029430 PRELIMINARY; PRT; 74 AA.
 AC 029430;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-AUG-1998 (TReMBLrel. 07, Last annotation update)
 DE HYPOTHETICAL 8.8 KDA PROTEIN.
 GN AF0828.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriplides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Retch C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL: AED01047; AAB90423.1; -;
 DR TIGR: AF0828; -;
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 74 AA; 8799 MW; 4DE93FB77E09CE56 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1;e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVK 6
 |||||
 Db 29 KRVK 33

RESULT 13
 ID 09GY26 PRELIMINARY; PRT; 82 AA.
 AC 09GY26;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE SM-G.
 GN SM-G.
 OS Trypanosoma brucei.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=427;
 RX MEDLINE=20381320; PubMed=10900267;
 RA Palfi Z., Tuecke S., Lahm H.-W., Lane W.S., Krut V.,
 RA Bragado-Nilsson E., Seraphin B., Blunderelf A.;
 RT "The spliceosomal snRNP core complex of Trypanosoma brucei: Cloning
 RT and functional analysis reveals seven Sm protein constituents.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8967-8972(2000).
 DR EMBL: AF280392; AAG00460.1; -;
 DR InterPro: IPR001163; snRNP_Sm.
 DR Pfam: PF01423; Sm; 1.
 SQ SEQUENCE 82 AA; 8922 MW; CF532BE52DD5F3A5 CRC64;

Query Match 83.3%; Score 5; DB 5; Length 82;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRVK 5
 |||||
 Db 15 EKRVK 19

RESULT 14
 ID 048711 PRELIMINARY; PRT; 88 AA.
 AC 048711;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
 DE SPAN GENE ENCODING NISIN AND INSERTION SEQUENCE IS904 (FRAGMENT).
 OS Lactococcus lactis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OX NCBI_TaxID=1358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Araya T.;
 RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D00696; BAA20961.1; -;
 FT NON_TER 88
 FT 88
 SQ SEQUENCE 88 AA; 10518 MW; 88D4CEB515622E44 CRC64;

Query Match 83.3%; Score 5; DB 2; Length 88;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVK 6
 |||||
 Db 68 KRVK 72

RESULT 15
 ID P70506 PRELIMINARY; PRT; 89 AA.
 AC P70506;
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-JAN-1999 (TReMBLrel. 09, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE GHF-1 PROTEIN.
 GN GHF-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PITUITARY TUMOUR;
 RX MEDLINE=92289691; PubMed-1600947;
 RA Theil L.E., Hattori K., Iazzaro D., Castriello J.L., Karin M.;
 RT "Differential splicing of the GHF1 primary transcript gives rise to
 two functionally distinct homeodomain proteins.";
 RL EMBO J. 11:2261-2269(1992).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER HOMEODOMAIN DOMAINS.
 DR EMBL: X65368; CAA46442.1; -.
 DR HSSP: p10037; 1A07.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 89 AA; 10661 MW; DD1BD16B7927589C CRC64;

Query Match 83.3%; Score 5; DB 11; Length 89;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKRVK 5
 [11111]
 Db 67 EKRVK 71

Search completed: February 26, 2002, 01:53:18
 Job time: 442 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:51:46 ; Search time 45.36 Seconds
(without alignments)
10.076 Million cell updates/sec

Title: US-09-819-136-2_COPY_283_288

Perfect score: 6

Sequence: 1 REPAR6

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : PIR.68:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	309	2	C72368
2	5	83.3	81	2	F83787
3	5	83.3	99	2	T36144
4	5	83.3	106	2	T30396
5	5	83.3	136	2	G71134
6	5	83.3	156	2	T49529
7	5	83.3	157	2	T08157
8	5	83.3	165	2	T41103
9	5	83.3	170	2	T28696
10	5	83.3	207	2	T37462
11	5	83.3	216	2	S32864
12	5	83.3	260	2	S71315
13	5	83.3	267	1	CTPGP
14	5	83.3	277	2	UC6142
15	5	83.3	277	2	E86659
16	5	83.3	283	2	D83863
17	5	83.3	289	2	S35300
18	5	83.3	294	2	H83544
19	5	83.3	305	2	F84982
20	5	83.3	319	2	F72750
21	5	83.3	344	2	E83364
22	5	83.3	344	2	G69068
23	5	83.3	381	2	T22008
24	5	83.3	387	2	H84209
25	5	83.3	393	2	A55859
26	5	83.3	436	2	UC1497
27	5	83.3	455	1	CHCHAI
28	5	83.3	455	2	J00158
29	5	83.3	455	2	A39062

30	5	83.3	456	2	A60652	gamma-aminobutyric
31	5	83.3	456	2	A27142	gamma-aminobutyric
32	5	83.3	457	2	F83064	probable MFS trans
33	5	83.3	468	2	E83046	seryl-tRNA(Sec) se
34	5	83.3	469	2	T08310	helicase homolog H
35	5	83.3	481	2	C71079	hypothetical prote
36	5	83.3	503	2	G75262	hypothetical prote
37	5	83.3	525	2	F70068	gamma-glutamyltran
38	5	83.3	529	2	T45134	hypothetical prote
39	5	83.3	537	2	T28683	hypothetical prote
40	5	83.3	545	2	H83342	periplasmic trehal
41	5	83.3	567	2	A33974	membrane transport
42	5	83.3	568	2	J03206	Uia6h protein - Ma
43	5	83.3	583	2	S29561	Ref(2)pp protein -
44	5	83.3	588	2	T35549	hypothetical prote
45	5	83.3	590	2	S29564	ref(2)ph protein -

ALIGNMENTS

RESULT 1
C72368
hypothetical protein TM0509 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: C72368
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; M0ID:9287316
A:Accession: C72368
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <ARN>
A:Cross-references: GB:AE001727; GB:AE000512; NID:g4981015; PIDN:AAD35594.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0509
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo
C.M.
Query Match 100.0%; Score 6; DB 2; Length 309;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 84 REPAR6 89
OY 1 REPAR6 6
111111
NO
RESULT 2
F83787
hypothetical protein BH1102 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: F83787
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; M0ID:20263314
A:Accession: F83787
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-81 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04821.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1102

Query Match 83.3%; Score 5; DB 2; Length 81;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 REPAIR 5
Db 66 REPAIR 70

RESULT 3

T36144
hypothetical protein SCE19A.21 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36144
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21598
A:Accession: T36144
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-99 <SEED>
A:Cross-references: EMBL:AL096852; PIDN:CAB51002.1; GSPDB:GN00070; SCOEDB:SCE19A.21
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE19A.21

Query Match 83.3%; Score 5; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAIR 5
Db 75 REPAIR 79

RESULT 4

T30396
hypothetical protein ORF49 - Lymantria dispar nuclear polyhedrosis virus

C:Species: Lymantria dispar nuclear polyhedrosis virus, LdNPV
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30396
R:Knizlo, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohlf
Virology 253, 17-34, 1999
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d
A:Reference number: Z20836; MUID:99124785
A:Accession: T30396
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-106 <KUZ>
A:Cross-references: EMBL:AF081810; NID:93822234; PIDN:MAC70234.1; PID:93822283

Query Match 83.3%; Score 5; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAIR 5
Db 87 REPAIR 91

RESULT 5

G71134
hypothetical protein PH0843 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: G71134
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137

A:Accession: G71134

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-136 <KAW>

A:Cross-references: GB:AP000003; NID:93236130; PIDN:BAA29937.1; PID:dl030880; PID:932

A:Experimental source: strain OT3

A>Note: this accession replaces an interim accession for a sequence replaced by Genba

C:Genetics:
A:Gene: PH0843

Query Match 83.3%; Score 5; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAIR 5
Db 52 REPAIR 56

RESULT 6

T49529
hypothetical protein B21J21.80 [imported] - Neurospora crassa

C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49529
R:Schulte, U.; Aign, V.; Heisele, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49529
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <SCH>
A:Cross-references: EMBL:AL355929; GSPDB:GN00116; NCSP:B21J21.80
A:Experimental source: BAC clone B21J21; strain OR74A
C:Genetics:
A:Gene: NCSP:B21J21.80
A:Map position: 6
C:Superfamily: Neurospora crassa hypothetical protein B21J21.80

Query Match 83.3%; Score 5; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 REPAIR 6
Db 86 REPAIR 90

RESULT 7

T08157
ribosomal protein L12 - Chlamydomonas reinhardtii (fragment)

C:Species: Chlamydomonas reinhardtii
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 20-Jun-2000
C:Accession: T08157
R:Walter, F.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z16385
A:Accession: T08157
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-157 <MAL>
A:Cross-references: EMBL:X95314; NID:q1167827; PIDN:CAA64626.1
C:Experimental source: strain 7781; clone 3
C:Superfamily: rat ribosomal protein L12
C:Keywords: protein biosynthesis; ribosome

Query Match 83.3%; Score 5; DB 2; Length 157;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
 |||||
 Db 79 EPARD 83

RESULT 8

T41103
 60S ribosomal protein l12 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
 C:Accession: T41103; T41292
 R:Purrelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21964
 A:Accession: T41103
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-165 <PUR>
 A:Cross-references: EMBL:AL031535; PIDN:CAA20752.1; GSPDB:GN00068; SPDB:SPCC16C4.13C
 A:Experimental source: strain 972h-; cosmid c16C4
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21985
 A:Accession: T41292
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-165 <MOO>
 A:Cross-references: EMBL:AL031824; PIDN:CAA21221.1; GSPDB:GN00068; SPDB:SPCC1H12.04C
 A:Experimental source: strain 972h-; cosmid c31H12
 C:Genetics:
 A:Gene: SPCC16C4.13C; SPCC1H12.04C
 A:Map position: 3
 A:introns: 4/3
 C:Superfamily: rat ribosomal protein l12

Query Match 83.3%; Score 5; DB 2; Length 165;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
 |||||
 Db 87 EPARD 91

RESULT 9

T28696
 hypothetical protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T28696
 R:Parthill, J.; Bentley, S.D.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, January 1999
 A:Reference number: Z20512
 A:Accession: T28696
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-170 <PAR>
 A:Cross-references: EMBL:AL023496; NID:e1292348; PID:e1292363; PIDN:CAA18913.1.

Query Match 83.3%; Score 5; DB 2; Length 170;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 Db 49 REPAR 53

RESULT 10

T37462

probable glutathione transferase (EC 2.5.1.18) - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T37462; T23483
 R:Tawe, W.N.; Eschbach, M.L.; Walter, R.D.; Henkle-Duehrsens, K.
 submitted to the EMBL Data Library, June 1997
 A:Description: Paraquat mediates differential gene expression in C. elegans.
 A:Reference number: Z21702
 A:Accession: T37462
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-207 <TAM>
 A:Cross-references: EMBL:AF010239; PIDN:AA65417.1
 A:Experimental source: strain Bristol N2
 R:Hendry, C.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: Z19746
 A:Accession: T23483

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-207 <WIL>
 A:Cross-references: EMBL:268879; PIDN:CAA93086.1; GSPDB:GN00022; CESP:K08F4.7
 A:Experimental source: clone K08F4
 C:Genetics:
 A:Gene: GST1; K08F4.7
 A:Map position: 4
 A:introns: 45/1; 76/1; 118/3
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 83.3%; Score 5; DB 2; Length 207;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
 |||||
 Db 127 EPARD 131

RESULT 11

S32864
 out1 protein - Erwinia carotovora
 C:Species: Erwinia carotovora
 C>Date: 08-Dec-1993 #sequence_revision 26-May-1995 #text_change 08-Oct-1999
 C:Accession: S32864; S31753
 R:Reeves, P.J.; Whitcombe, D.; Wharam, S.; Gibson, M.; Allison, G.; Bunce, N.; Barrell, M.L.
 Microbiol. 8, 443-456, 1993
 A:Title: Molecular cloning and characterization of 13 out genes from Erwinia carotovora bacteria.
 A:Reference number: S32857; NID:93316842
 A:Accession: S32864
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-216 <REE>
 A:Cross-references: EMBL:X70049; NID:g42184; PIDN:CAA9651.1; PID:g42192
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, //.
 C:Genetics:
 A:Gene: out1
 C:Superfamily: secretion protein xcpW

Query Match 83.3%; Score 5; DB 2; Length 216;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 Db 15 REPAR 19

RESULT 12

S71315

deoxyguanosine kinase (EC 2.7.1.113) precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
 C:Accession: S71315; S78432
 R:Wang, L.; Hellman, U.; Eriksson, S.
 FEBS Lett. 390, 39-43, 1996
 A:Title: Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA.
 A:Reference number: S71315; MUID:96314545
 A:Accession: S71315
 A:Molecule type: mRNA
 A:Residues: 1-260 <MAN>
 A:Cross-references: EMBL:X87386; PIDN:CAA6054.1
 A:Experimental source: tissue brain
 R:Wang, L.; Hellman, U.; Eriksson, S.
 Submitted to the EMBL Data Library, April 1996
 A:Description: Cloning and expression of human deoxyguanosine kinase cDNA.
 A:Reference number: S78432
 A:Accession: S78432
 A:Molecule type: mRNA
 A:Residues: 1-18, 'R', 20-260 <MAN>
 A:Cross-references: EMBL:X97386; PIDN:CAA6054.1
 A:Experimental source: brain
 C:Genetics:
 A:Gene: DGK
 A:Genome: nuclear
 C:Superfamily: human deoxycytidine kinase
 C:Keywords: mitochondrion; phosphotransferase
 F:1-22/Domains: transit peptide (mitochondrion) #status predicted <TNP>
 F:23-260/Product: deoxyguanosine kinase #status predicted <MAT>
 F:125-132/Region: DRS motif
 F:185-191/Region: arginine-rich

Query Match 83.3%; Score 5; DB 2; Length 260;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 REPAR 5
 11111
 Db 84 REPAR 88

RESULT 13
 CTPGP
 corticotropin / lipotropin precursor - pig
 N:Alternate names: POMC; proopiomelanocortin
 C:Contains: beta-endorphin; CLIP; corticotropin; lipotropin beta; melanotropin alpha; me
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 31-Dec-1979 #sequence_revision 30-Jun-1987 #text_change 21-Jan-2000
 C:Accession: A93496; A90657; A93383; A91779; A93141; A90580; A93140; A91782; A93177; A90
 R:Boileau, G.; Barbeau, C.; Jeanmoult, L.; Chretien, M.; Drouin, J.
 Nucleic Acids Res. 11, 8063-8071, 1983
 A:Title: Complete structure of the porcine pro-opiomelanocortin mRNA derived from the nu
 A:Reference number: A93496; MUID:84069823
 A:Accession: A93496
 A:Molecule type: mRNA
 A:Residues: 1-267 <BOI>
 A:Cross-references: GB:X00135; GB:X01879; NID:g2070; PIDN:CAA24968.1; PID:g758337
 R:Gossard, F.J.; Chang, A.C.Y.; Cohen, S.N.
 Biochim. Biophys. Acta 866, 68-74, 1986
 A:Title: Sequence of the cDNA encoding porcine pro-opiomelanocortin.
 A:Reference number: A90657; MUID:86131687
 A:Accession: A90657
 A:Molecule type: mRNA
 A:Residues: 1-5, 'S', 7-14, 'A', 16-22, 'E', 24-267 <GOS>
 R:Schally, A.V.; Guoch, J.G.; Redding, T.W.; Groot, K.; Rodriguez, H.; Szonyi, E.; Stuit
 Proc. Natl. Acad. Sci. U.S.A. 88, 3540-3544, 1991
 A:Title: Isolation and characterization of two peptides with prolactin release-inhibiti
 A:Reference number: A93383; MUID:91219407
 A:Accession: A93383
 A:Molecule type: protein
 A:Residues: 27-52 <SCH>
 R:Shepherd, R.G.; Willson, S.D.; Howard, K.S.; Bell, P.H.; Davies, D.S.; Davis, S.B.; El

J. Am. Chem. Soc. 78, 5067-5076, 1956
 A:Title: Studies with corticotropin. III. Determination of the structure of beta-cort
 A:Reference number: A91779
 A:Accession: A91779
 A:Molecule type: protein
 A:Residues: 136-174 <SHE>
 R:Piniker, B.; Steber, P.; Rittel, W.; Zuber, H.
 Nature New Biol. 235, 114-115, 1972
 A:Title: Revised amino-acid sequences for porcine and human adrenocorticotrophic horm
 A:Reference number: A93403; MUID:72114902
 A:Contents: annotation; revision to residues 160 and 165
 R:Harris, J.I.; Lerner, A.B.
 Nature 179, 1346-1347, 1957
 A:Title: Amino acid sequence of the alpha-melanocyte-stimulating hormone.
 A:Reference number: A93141
 A:Accession: A93141
 A:Molecule type: protein
 A:Residues: 136-148 <HAR>
 R:Graf, L.; Baratt, E.; Casch, G.; Sajgo, M.
 Biochim. Biophys. Acta 229, 276-278, 1971
 A:Title: Amino acid sequence of porcine beta-lipotrophic hormone.
 A:Reference number: A90580; MUID:71111231
 A:Accession: A90580
 A:Molecule type: protein
 A:Residues: 154-174 <GR1>
 R:Gillardeau, C.; Chretien, M.
 in Chemistry and Biology of Peptides, Melenhof, J., ed., pp.609-611, Ann Arbor Sci.
 A:Title: Complete amino acid sequence of porcine beta-lipotrophic hormone (beta-LPH).
 A:Reference number: A94412
 A:Contents: annotation; revised sequence of lipotropin
 R:Pankov, Y.A.; Yudaev, N.A.
 Biochimica 37, 991-1004, 1972
 A:Reference number: A90679; MUID:73048217
 A:Contents: annotation; revision to residue 211
 R:Harris, J.I.; Roos, P.
 Nature 178, 90, 1956
 A:Title: Amino-acid sequence of a melanophore-stimulating peptide.
 A:Reference number: A93140
 A:Accession: A93140
 A:Molecule type: protein
 A:Residues: 217-234 <HAR2>
 R:Gschwind, I.I.; Li, C.H.; Barnafl, L.
 J. Am. Chem. Soc. 79, 620-625, 1957
 A:Title: The structure of the beta-melanocyte-stimulating hormone.
 A:Reference number: A91782
 A:Accession: A91782
 A:Molecule type: protein
 A:Residues: 217-234 <GES>
 R:Hughes, J.; Smith, T.W.; Kosterlitz, H.W.; Fothergill, L.A.; Morgan, B.A.; Morris,
 Nature 258, 577-579, 1975
 A:Title: Identification of two related pentapeptides from the brain with potent opiat
 A:Reference number: A93177; MUID:76100762
 A:Accession: A93177
 A:Molecule type: protein
 A:Residues: 237-241 <HUG>
 R:Graf, L.; Baratt, E.; Patthy, A.
 Acta Biochim. Biophys. Acad. Sci. Hung. 11, 121-122, 1976
 A:Title: Isolation of a COOH-terminal beta-lipotropin fragment (residues 61-91) with
 A:Reference number: A90001; MUID:77084500
 A:Accession: A90001
 A:Molecule type: protein
 A:Residues: 237-267 <GR2>
 R:Voigt, K.; Stegmayer, W.; McGregor, G.P.; Roesch, H.; Seliger, H.
 Eur. J. Biochem. 194, 225-236, 1990
 A:Title: Isolation and full structural characterisation of six adrenocorticotropin-11
 o forms of a novel adrenocorticotropin-like peptide.
 A:Reference number: S13845; MUID:91071194
 A:Accession: S13845
 A:Molecule type: protein
 A:Residues: 136-165, 'S', 167-174 <VOI>
 A:Experimental source: pituitary
 C:Superfamily: corticotropin-lipotropin
 C:Keywords: glycoprotein; hormone; neuropeptide; opioid peptide; pituitary

F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-52/Product: prolactin release-inhibiting factor 1 #status experimental <PRIL>
 F:136-174/Product: corticotropin #status experimental <CTR>
 F:136-148/Product: melanotropin alpha #status experimental <MTA>
 F:154-114/Product: CLIP #status experimental <CLP>
 F:177-267/Product: lipotropin beta #status predicted <LTB>
 F:217-234/Product: melanotropin beta #status experimental <MTB>
 F:237-267/Product: beta-endorphin #status experimental <BEF>
 F:237-241/Product: Met-enkephalin #status experimental <MEN>
 F:1/Binding site: carbonyl site (Thr) (covalent) #status experimental
 F:91/Binding site: carbonyl site (Asn) (covalent) #status experimental

Query Match 83.3%; Score 5; DB 1; length 267;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
 |||||
 DB 184 EPARD 188

RESULT 14
 JC6142
 deoxyguanosine kinase (EC 2.7.1.113) precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 18-Jun-1999
 C:Accession: JC6142
 R:Johansson, M.; Karlsson, A.
 Proc. Natl. Acad. Sci. U.S.A. 93, 7258-7262, 1996
 A>Title: Cloning and expression of human deoxyguanosine kinase cDNA.
 A:Reference number: JC6142; MUID:96293511
 A:Accession: JC6142
 A:Molecule type: mRNA
 A:Residues: 1-277 <JOH>
 A:Cross-references: GB:U41668; NID:q1477481; PIDN:AAC50624.1; PID:q1477482
 A>Note: It is uncertain whether Met-1 or Met-8 is the initiator
 C:Comment: This enzyme is involved in mediating cytotoxicity of nucleoside analogs. It F
 inase, thymidine kinase 2.
 C:Superfamily: human deoxycytidine kinase
 C:Keywords: phosphotransferase
 F:1-39/Domain: signal sequence #status predicted <SIG>
 F:40-277/Product: deoxyguanosine kinase #status predicted <MAT>

Query Match 83.3%; Score 5; DB 2; length 277;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 DB 101 REPAR 105

RESULT 15
 EB6659
 ABC transporter ATP-binding protein ychd [imported] - Lactococcus lactis subsp. lactis
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
 C:Accession: EB6659
 R:Bohlool, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich
 Genome Res. In press, 2001
 A>Title: The complete genome sequence of the lactic acid bacterium.
 A:Reference number: A86625
 A:Accession: EB6659
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-277 <STO>
 A:Cross-references: GB:AE005176; NID:q12723139; PIDN:AAK04375.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: ychd

Query Match 83.3%; Score 5; DB 2; length 277;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 DB 136 REPAR 140

Search completed: February 26, 2002, 01:51:47
 Job time: 456 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:53:18 ; Search time 78.18 seconds

(without alignments)
11.226 Million cell updates/sec

Title: US-09-819-136-2_COPY_283_288

Perfect score: 6

Sequence: 1 REPARD 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size: 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SPTREMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_fodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.0	309	2	Q9WYX9	O9WYX9 thermotoga
2	100.0	527	2	Q9RD76	O9RD76 streptomyc
3	100.0	556	2	Q460A3	Q460A3 cellulosoma
4	83.3	81	2	Q9KDV9	Q9KDV9 bacillus ha
5	83.3	99	2	Q9S2G0	Q9S2G0 streptomyc
6	83.3	106	12	Q9YMS5	Q9YMS5 lymantria d
7	83.3	110	5	Q9ND17	Q9ND17 jaas edwar
8	83.3	113	5	Q9ND16	Q9ND16 jaas edwar
9	83.3	124	4	Q9BXC7	Q9BXC7 homo sapien
10	83.3	132	2	Q9L1S8	Q9L1S8 streptomyc
11	83.3	136	1	Q58573	Q58573 pyrococcus
12	83.3	136	2	Q9XC36	Q9XC36 pseudomonas
13	83.3	154	2	Q9A562	Q9A562 caulobacter
14	83.3	156	3	Q9P6A0	Q9P6A0 neurospora
15	83.3	168	12	Q9DH63	Q9DH63 meleagrid h
16	83.3	169	11	Q55315	Q55315 mus musculu
17	83.3	170	2	Q69820	Q69820 streptomyc
18	83.3	184	12	Q9DRS7	Q9DRS7 human immun
19	83.3	206	12	Q93010	Q93010 human immun

20	83.3	207	5	Q21355	Q21355 caenorhabdit
21	83.3	221	4	Q9C029	Q9C029 homo sapien
22	83.3	233	2	Q9F2R8	Q9F2R8 streptomyc
23	83.3	243	11	Q99PQ5	Q99PQ5 mus musculu
24	83.3	254	5	Q9VF16	Q9VF16 drosophila
25	83.3	270	3	Q9HEF6	Q9HEF6 neurospora
26	83.3	274	5	Q9VNY3	Q9VNY3 drosophila
27	83.3	277	2	Q9C1S9	Q9C1S9 lactococcus
28	83.3	279	2	Q9L789	Q9L789 alcalligenes
29	83.3	281	2	Q9A672	Q9A672 caulobacter
30	83.3	282	2	Q9KC66	Q9KC66 bacillus ha
31	83.3	289	2	Q56905	Q56905 yersinia en
32	83.3	294	2	Q915C3	Q915C3 pseudomonas
33	83.3	302	12	Q9DM7	Q9DM7 rat cytomeg
34	83.3	307	5	Q9MKD3	Q9MKD3 drosophila
35	83.3	313	10	Q9M8S7	Q9M8S7 arabidopsis
36	83.3	319	1	Q9YEP8	Q9YEP8 aeropyrum p
37	83.3	327	12	Q9DH84	Q9DH84 gallid hep
38	83.3	338	2	Q9K303	Q9K303 streptomyc
39	83.3	340	2	Q84915	Q84915 pseudomonas
40	83.3	344	1	Q27556	Q27556 methanobact
41	83.3	376	10	Q9ST77	Q9ST77 oryza sativ
42	83.3	381	5	Q18161	Q18161 caenorhabdit
43	83.3	387	1	Q9HRW4	Q9HRW4 halobacteri
44	83.3	396	2	Q48468	Q48468 klebsiella
45	83.3	399	10	Q91162	Q91162 arabidopsis

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	309 AA.
Q9WYX9	Q9WYX9			
AC	Q9WYX9			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	UDP-GLUCOSE 4-EPIMERASE, PUTATIVE.			
GN	TM0509.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogales; Thermotoga.			
OX	NCBI_TaxID=2336;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MSB8 / DSM 3109;			
RC	MEDLINE=99287316; PubMed=10360571;			
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,			
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,			
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,			
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,			
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,			
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,			
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from			
RT	genome sequence of Thermotoga maritima."			
RL	Nature 399:323-329 (1999).			
EMBL	AE001727; AAD35594.1; -			
DR	HSSP: P09147; IYEL.			
DR	TIGR: TM0509; -			
DR	InterPro: IPR002198; ADH_short.			
DR	InterPro: IPR001509; Epimerase.			
DR	InterPro: IPR000205; NAD-binding.			
DR	InterPro: IPR000594; TH1F-family.			
DR	Pfam: PF01370; Epimerase; 1.			
DR	PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.			
KW	Complete proteome.			
SC	SEQUENCE 309 AA; 34900 MW; 0200114370FID001 CRC64;			

Query Match 100.0%; Score 6; DB 2; Length 309;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 6
 DB 84 REPAR 89

RESULT 2

OY 09RD76 PRELIMINARY; PRT; 527 AA.
 AC 09RD76; 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PUTATIVE CYTOCHROME P450.
 GN SCF43.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA MEDLINE-97000351; PubMed-8843436;
 RX Kadenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
 BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
 CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: A136502; CAB6201.1; -.
 DR HSP: P14779; 1B07.
 DR InterPro: IPR001128; Cyt_P450.
 DR InterPro: IPR000217; Tubulin.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW GTP-binding; Heme; Microtubules; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 527 AA; 57741 MW; 4A612C1B7A894780 CRC64;

Query Match 100.0%; Score 6; DB 2; Length 527;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 6
 DB 6 REPAR 11

NO

RESULT 3
 ID 046043 PRELIMINARY; PRT; 556 AA.
 AC 046043; 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE BETA-GLUCOSIDASE.
 OS Cellulomonas fimi.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kim H.K., Pack M.Y.;
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL: M94865; AAA23091.1; -.
 DR HSP: P11546; 1PB6.
 DR InterPro: IPR001360; Glyco_hydro_1.
 DR Pfam: PF00232; Glyco_hydro_1; 2.
 DR PROSITE: PS00653; GLYCOSYL_HYDROL_F1_2; 1.
 SQ SEQUENCE 556 AA; 62527 MW; CC4706F72A7A96B3 CRC64;

Query Match 100.0%; Score 6; DB 2; Length 556;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 6
 DB 299 REPAR 304

NO

RESULT 4
 ID 09KDV9 PRELIMINARY; PRT; 81 AA.
 AC 09KDV9; 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE BH1102 PROTEIN.
 GN BH1102.
 OS Bacillus halodurans.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86655;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / UCM 9153;
 RX MEDLINE-20512582; PubMed-11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001510; BAB04821.1; -.
 KW Complete proteome.
 SQ SEQUENCE 81 AA; 9241 MW; 875FA0163C21DB3 CRC64;

Query Match 83.3%; Score 5; DB 2; Length 81;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 DB 66 REPAR 70

RESULT 5
 ID 09S2G0 PRELIMINARY; PRT; 99 AA.
 AC 09S2G0; 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE HYPOTHETICAL 10.0 KDA PROTEIN.
 GN SCE19A.21.
 OS Streptomyces coelicolor.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID-1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE-97000351; PubMed-8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL096852; CAB51002.1; -;
KW Hypothetical protein.
SQ SEQUENCE 99 AA; 10026 MW; EC53B35737CCA8C CRC64;

Query Match 83.3%; Score 5; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAIR 5
DB 75 REPAIR 79

RESULT 6
O9YMS5 PRELIMINARY; PRT; 106 AA.
ID O9YMS5;
AC O9YMS5;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE LDORF-49 PEPTIDE.
OS Lymantria dispar multicapsid nuclear polyhedrosis virus (LdMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID-10449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99124785; PubMed-9887315;
RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA Slavicek J.M., Rohmann G.F.;
RT "Sequence and analysis of the genome of a baculovirus pathogenic for
RT Lymantria dispar.";
RL Virology 253:17-34(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA Slavicek J., Rohmann G.F.;
RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kuzio J.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF081810; AAC70234.1; -;
SQ SEQUENCE 106 AA; 12487 MW; 88A368C458CA0B8 CRC64;

Query Match 83.3%; Score 5; DB 12; Length 106;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAIR 5
DB 11111

DB 87 REPAIR 91

RESULT 7
O9ND17 PRELIMINARY; PRT; 110 AA.
ID O9ND17;
AC O9ND17;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE EYESTALK PEPTIDE.
OS Jasus edwardsii (red rock lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;
OC Palinuroidea; Palinuridae; Jasus.
OX NCBI_TaxID-95461;
RN [1]
RP SEQUENCE FROM N.A.
RA Khoo J.G.I., Sin F.Y.T.;
RT "A novel eyestalk peptide from the lobster Jasus edwardsii.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF112985; AAF78450.1; -;
SQ SEQUENCE 110 AA; 11178 MW; E4D9538AB52B9CA5 CRC64;

Query Match 83.3%; Score 5; DB 5; Length 110;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAIR 5
DB 82 REPAIR 86

RESULT 8
O9ND16 PRELIMINARY; PRT; 113 AA.
ID O9ND16;
AC O9ND16;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE EYESTALK PEPTIDE.
OS Jasus edwardsii (red rock lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;
OC Palinuroidea; Palinuridae; Jasus.
OX NCBI_TaxID-95461;
RN [1]
RP SEQUENCE FROM N.A.
RA Khoo J.G.I., Sin F.Y.T.;
RT "A novel eyestalk peptide from the lobster Jasus edwardsii.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF112986; AAF78451.1; -;
SQ SEQUENCE 113 AA; 11603 MW; A09F1DE09B918B17 CRC64;

Query Match 83.3%; Score 5; DB 5; Length 113;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAIR 5
DB 85 REPAIR 89

RESULT 9
O9BX07 PRELIMINARY; PRT; 124 AA.
ID O9BX07;
AC O9BX07;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CAT EYE SYNDROME CRITICAL REGION CANDIDATE GENE NUMBER 4 (FRAGMENT).

GN CECR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Foote T., Brinkman-Mills P., Banting G.S., Maier S.A., Riazzi M.A.,
RA Bridgland L., Hu S., Birren B., Minoshima S., Shimizu N., Pan H.,
RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shauli S., Phan S., Yao Z.,
RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,
RA McDermid H.,
RT "Analysis of the Cat eye syndrome critical region candidate gene
RT number in humans and the region of conserved synteny in mice : A
RT search for candidate genes at or near the human chromosome 22
RT pericentromere."
RL Genome Res. 0:0-0(2001).
DR EMBL: AF307448; AAK30048.1; -.
PT NON_TER
SQ SEQUENCE 124 AA; 12838 MW; 938E00386308EC7A CRC64;

Query Match 83.3%; Score 5; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAIR 5
DB 101 REPAIR 105

RESULT 10
ID Q9L158 PRELIMINARY; PRT; 132 AA.
AC Q9L158;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 13.8 KDA PROTEIN.
GN SEC31.14C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K.J., Harris D.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Thompson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL158060; CAB76320.1; -.
KM Hypothetical protein.
SQ SEQUENCE 132 AA; 13791 MW; D5BBF63912D774B CRC64;

Query Match 83.3%; Score 5; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 EPARD 6

DB 11 EPARD 15
RESULT 11
ID 058573 PRELIMINARY; PRT; 136 AA.
AC 058573;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE HYPOTHETICAL 15.9 KDA PROTEIN PH0843.
GN PH0843.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamida M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000003; BAA29937.1; -.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 136 AA; 15930 MW; 39BAA61A52F5BFO CRC64;

Query Match 83.3%; Score 5; DB 1; Length 136;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAIR 5
DB 52 REPAIR 56

RESULT 12
ID Q9XC36 PRELIMINARY; PRT; 136 AA.
AC Q9XC36;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 15.6 KDA PROTEIN (PDTORFG).
GN PDTORFG.
OS Pseudomonas stutzeri.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KC;
RA Sepulveda-Torres L.C., Criddle C.S.;
RT "Sequence of a 8,273 bp DNA sequence from Pseudomonas sp. KC with
RT apparent involvement in carbon tetrachloride degradation."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KC;
RA Lewis T.A., Cortese M.S., Sebat J.L., Green T.L., Crawford R.L.;
RT "Identification of a Region of the Pseudomonas stutzeri strain KC
RT chromosome containing genes for the Biosynthesis of Pyridine-2,6-
RT bis(thiocarboxylic acid), the Agent of Carbon Tetrachloride
RT Dechlorination Produced by this Organism."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF149851; AAD39226.1; -.

DR EMBL: AF196567; AAF3131.1; -
 DR InterPro: IPR000555; Mov34.
 DR Pfam: PF01398; Mov34; 1.
 DR SMART: SM00232; JAB_MPN; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 136 AA; 15627 MW; 92B03CD09EF7F257 CRC64;

Query Match 83.3%; Score 5; DB 2; Length 136;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 DB 114 REPAR 118

RESULT 13
 O9A562 PRELIMINARY; PRT; 154 AA.
 AC O9A562;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE HYPOTHEICAL PROTEIN CC2602.
 GN CC2602.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21173698; PubMed-11259647;
 RA Eisen M.W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Nierman W.C., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Debey R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AE005928; AAK24571.1; -
 DR TIGR: CC2602; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 154 AA; 17218 MW; F7A0D2E1B9D3D3B CRC64;

Query Match 83.3%; Score 5; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 DB 11 REPAR 15

RESULT 14
 O9P640 PRELIMINARY; PRT; 156 AA.
 AC O9P640;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHEICAL 18.2 KDA PROTEIN.
 GN B21J21.80.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariatales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Algn V., Hohnsels J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RA Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL355929; CAB91330.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 156 AA; 18200 MW; 61AA30F00FC4AB23 CRC64;

Query Match 83.3%; Score 5; DB 3; Length 156;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
 |||||
 DB 86 EPARD 90

RESULT 15
 O9DH63 PRELIMINARY; PRT; 168 AA.
 AC O9DH63;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE HYPOTHEICAL 18.8 KDA PROTEIN.
 GN HVT075.
 OS Meleagrid herpesvirus 1 (herpesvirus of turkeys).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OX NCBI_TaxID=37108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FC126;
 RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Rock D.L., Kutish G.K.;
 RT "The genome of turkey herpesvirus."
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF291866; AAG45822.1; -
 DR EMBL: AF291866; AAG45734.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 168 AA; 18798 MW; 40A5E01D617DD2CD CRC64;

Query Match 83.3%; Score 5; DB 12; Length 168;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 DB 56 REPAR 60

Search completed: February 26, 2002, 01:53:19
 Job time: 443 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:53:53 ; Search time 24.25 Seconds
(without alignments)
9.072 Million cell updates/sec

Title: US-09-819-136-2_COPY_525_530

Perfect score: 6

Sequence: 1 EKRYKK 6

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	691	1	TOP1_BACSU
2	5	83.3	80	1	YF41_ARCFU
3	5	83.3	118	1	RS13_MYCSP
4	5	83.3	128	1	RS11_HAEIN
5	5	83.3	131	1	RS11_BUCAT
6	5	83.3	144	1	Y047_ARCFU
7	5	83.3	147	1	H2B_GOSHI
8	5	83.3	147	1	SSRP_MYCPN
9	5	83.3	151	1	UBC6_DROME
10	5	83.3	156	1	Y723_METUA
11	5	83.3	162	1	YR27_CAEEL
12	5	83.3	167	1	YF81_AQUAE
13	5	83.3	171	1	HSCB_ECOLI
14	5	83.3	181	1	RM06_ACACA
15	5	83.3	185	1	IF3C_CYACA
16	5	83.3	205	1	MUSG_SYNY3
17	5	83.3	219	1	RB3B_BOVIN
18	5	83.3	219	1	RB3B_HUMAN
19	5	83.3	219	1	RB3B_RAT
20	5	83.3	245	1	YD76_METUA
21	5	83.3	261	1	FABI_RICPR
22	5	83.3	271	1	RL7A_ANOGA
23	5	83.3	281	1	TRPA_METUA
24	5	83.3	286	1	DMA_HAEIN
25	5	83.3	291	1	PIR1_BOVIN
26	5	83.3	291	1	PIR1_CANFA
27	5	83.3	291	1	PIR1_HUMAN
28	5	83.3	291	1	PIR1_MACMU
29	5	83.3	291	1	PIR1_MOUSE
30	5	83.3	291	1	PIR1_PIG
31	5	83.3	291	1	PIR1_RAT
32	5	83.3	291	1	PIR1_SHEEP
33	5	83.3	294	1	MUSG_STRGR

ALIGNMENTS

RESULT	1	STANDARD	PRT	691 AA.
TOP1_BACSU				
ID	TOP1_BACSU			
AC	P39814			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)			
DE	(UNWINDING ENZYME) (SWIHELASE).			
OS	TOPA OR TOP1.			
GN	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168 / 865;			
RA	de Jong S.;			
RT	"Cloning and sequencing of the Top1 gene, the gene encoding B. subtilis DNA topoisomerase I."			
RL	Submitted (XXX-1994) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RA	Fouquier D., Errington J.;			
RT	"Cloning and sequencing 7.5kbp of DNA from Bacillus subtilis upstream of the cody gene."			
RL	Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.			
CC	-1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED DNA, FOLLOWED BY PASSAGE AND REJOINING.			
CC	-1- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS AT ONE END OF THE ENZYME-SEVERED DNA STRAND.			
CC	-1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: L27797; AAA22763.1; -			
CC	EMBL: AJ000975; CAA04422.1; -			
CC	EMBL: Z99112; CAB13485.1; -			
CC	HSSP: P06612; IECL.			
CC	Subtilist; BG11007; topA.			
CC	InterPro: IPR003601; DNATopI_ABP_bind.			
CC	DR InterPro: IPR003602; DNATopI_DNA_bind.			
CC	InterPro: IPR000380; Pro_topoisomre.			

34	5	83.3	295	1	GALT_HAEIN	P44878 haemophilus
35	5	83.3	295	1	YF67_ARCFU	O28705 archaeoglob
36	5	83.3	298	1	SNAB_BOVIN	P81126 bos taurus
37	5	83.3	299	1	MUSG_STRVG	P27309 streptomyc
38	5	83.3	301	1	GALT_ECOLI	P25520 escherichia
39	5	83.3	306	1	GTRB_ECOLI	P77293 escherichia
40	5	83.3	306	1	YDPC_BACSU	P66680 bacillus su
41	5	83.3	317	1	Y302_MYCGE	P47544 mycoplasma
42	5	83.3	327	1	YAB6_SCHPO	O09713 schizosacch
43	5	83.3	332	1	LIPB_KLULA	O13476 kluyveromyc
44	5	83.3	342	1	QUEA_BACSU	O32054 bacillus su
45	5	83.3	347	1	FPPS_SCHPO	O14230 schizosacch

DR InterPro: IPR002936; Toprim.
DR Pfam: PF01131; Topoisom_bac; 1.
DR Pfam: PF01751; Toprim; 1.
DR Pfam: PF01396; zf-C4_Topoisom; 3.
DR PRINTS: PR00417; PRPISMBASE1.
DR SMART: SM00437; TOP1AC; 1.
DR SMART: SM00436; TOP1BC; 1.
DR SMART: SM00493; TOP1M; 1.
DR PROSITE: PS00396; TOPOISOMERASE_I_PROK; 1.
DR Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
KW Repeat: Complete proteome.
FT ZN_FING 579 605 C4-TYPE 1.
FT ZN_FING 619 647 C4-TYPE 2.
FT ZN_FING 660 683 C4-TYPE 3.
FT ACT_SITE 298 DNA_CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 691 AA; 79078 MW; 58A45C84E52711BB CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 691;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKVK 6
|11111|
Db 555 EKKVK 560

RESULT 2
YJ41_ARCFU STANDARD; PRT; 80 AA.
ID YJ41_ARCFU
AC 028338;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AF1941.
GN AF1941.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
CC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kechum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriakides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kinkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöck A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: BELONGS TO THE UPF0062 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000969; AAB89310.1; -
DR TIGR: AF1941; -
DR InterPro: IPR003850; UPF0062.
DR Pfam: PF02700; UPF0062; 1.
KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 80 AA; 9221 MW; BF28F8BF23E4DE78 CRC64;

Query Match
Best Local Similarity 100.0%; Score 5; DB 1; Length 80;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRVKK 6
|11111|
Db 31 KRVKK 35

RESULT 3
RS13_MYCSP STANDARD; PRT; 118 AA.
ID RS13_MYCSP
AC P38017;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S13.
GN RPSM.
OS Mycoplasma sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94042887; PubMed=8226662;
RA Tan M., Klein R., Grant R., Ganem D., Engel J.N.;
RT "Cloning and characterization of the RNA polymerase alpha-subunit
J. Bacteriol. 175:7150-7159(1993).
[2]
RP CORRECTION OF SPECIES OF ORIGIN.
RA Tan M., Klein R., Grant R., Ganem D., Engel J.N.;
RL J. Bacteriol. 177:2607-2607(1995).
CC -1- FUNCTION: INVOLVED IN THE BINDING OF FMET-tRNA AND, HENCE, IN THE
CC INITIATION OF TRANSLATION.
CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO ORIGINATE FROM
CC CHLAMYDIA TRACHOMATIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L23478; AAA16205.1; -
DR PIR: C49923; C49923.
DR InterPro: IPR001892; Ribosomal_S13.
DR Pfam: PF00416; Ribosomal_S13; 1.
DR Prodom: PP001363; Ribosomal_S13; 1.
DR PROSITE: PS00646; RIBOSOMAL_S13; FALSF_NEC.
KW RIBOSOMAL PROTEIN.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 118 AA; 13919 MW; A115CEBF078AD11C CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 118;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKVK 5
|11111|
Db 41 EKKVK 45

RESULT 4
RS11_HAEIN STANDARD; PRT; 128 AA.
ID RS11_HAEIN

AC P44379;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE 30S RIBOSOMAL PROTEIN S11.
 GN RPSK OR RPS11 OR H10800.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 CC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Uettermann J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanne M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Furmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: S11 PLAYS AN ESSENTIAL ROLE FOR THE SELECTION OF THE
 CC CORRECT tRNA IN PROTEIN BIOSYNTHESIS. IT IS LOCATED ON THE LARGE
 CC LOBE OF THE SMALL SUBUNIT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S11P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U32762; AAC2459.1; -
 CC TIGR: H10800; -
 DR InterPro: IPR001971; Ribosomal_S11.
 DR Pfam: PF00411; Ribosomal_S11; 1.
 DR ProDom: PD001010; Ribosomal_S11; 1.
 DR PROSITE: PS00054; RIBOSOMAL_S11; 1.
 KW Ribosomal protein; Complete proteome.
 FT INIT_MER 0
 FT BY SIMILARITY.
 SQ SEQUENCE 128 AA; 13853 MW; CC71D066748D6A70 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6
 Db 9 KRVKK 13
 ID RSL1_BUCAT
 ID RSL1_BUCAT STANDARD; PRT; 131 AA.
 AC P57568;
 DT 20-AUG-2001 (rel. 40, Created)
 DT 20-AUG-2001 (rel. 40, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE 30S RIBOSOMAL PROTEIN S11.
 GN RPSK OR B0501.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 CC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 CC NCBI_TaxID=118099;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=TOXKO 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: S11 PLAYS AN ESSENTIAL ROLE FOR THE SELECTION OF THE
 CC CORRECT tRNA IN PROTEIN BIOSYNTHESIS. IT IS LOCATED ON THE LARGE
 CC LOBE OF THE SMALL SUBUNIT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S11P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF001119; BAB1394.1; -
 CC InterPro: IPR001971; Ribosomal_S11.
 DR Pfam: PF00411; Ribosomal_S11; 1.
 DR ProDom: PD001010; Ribosomal_S11; 1.
 DR PROSITE: PS00054; RIBOSOMAL_S11; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 131 AA; 14208 MW; 82B3A77EE42ACD38 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 131;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6
 Db 12 KRVKK 16
 ID Y047_ARCFU
 ID Y047_ARCFU STANDARD; PRT; 144 AA.
 AC O30189;
 DT 20-AUG-2001 (rel. 40, Created)
 DT 20-AUG-2001 (rel. 40, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AF0047.
 GN AF0047.
 OS Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 CC NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulfate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF001103; AAB91185.1; -
 DR TIGR: AF0047; -
 KW Hypothetical protein; Coiled coil; Complete proteome.
 FT DOMAIN 48 119 COILED COIL (POTENTIAL).
 SQ SEQUENCE 144 AA: 16896 MW: 219936D208E250C CRC64;

Query Match 83.3%; Score 5; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRVK 5
 DB 24 KRVK 28

RESULT 7
 H2B_GOSHI STANDARD; PRT; 147 AA.
 AC 022582;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HISTONE H2B.
 GN HIS2B.
 OS Gossypium hirsutum (Upland cotton).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Malvales; Malvaceae; Gossypium.
 CC NCBI_TaxID=3635;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=CV. DELTAPINE 62; TISSUE=Etisolated cotyledon;
 CC RA Tuley R.B.;
 CC "cDNA clones encoding histone H3 and histone H2B from upland cotton
 CC (Gossypium hirsutum L.)."
 CC RL (in) Plant Gene Register GGR97-182.
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
 CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF025667; AAB91163.1; -
 DR Mende1; 26413; Goshi.HIS2B;26413.
 DR InterPro: IPR000558; Histone_H2B.
 DR InterPro: IPR000166; Histone_core.
 DR Pfam: PF00125; histone; 1.
 DR PRINTS: PR00621; HISTONEH2B.
 DR PRODOM: PD000497; Histone_H2B; 1.
 DR SMART: SM00427; H2B; 1.
 DR SMART: SM00357; HISTONE_H2B; 1.
 DR PROSITE: PS00357; HISTONE_H2B; 1.
 KW Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding.
 SQ SEQUENCE 147 AA: 16087 MW: 66F5C774E6E11F6 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 KRVK 6

DB 51 KRVK 55

RESULT 8
 SSRP_MYCPN STANDARD; PRT; 147 AA.
 ID SSRP_MYCPN
 AC P75043;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SSRP-BINDING PROTEIN.
 GN SMPB OR MPN074 OR MP081.
 OS Mycoplasma pneumoniae.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma.
 CC NCBI_TaxID=2104;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=ATCC 29342 / M129;
 CC MEDLINE=97105885; PubMed=8948633;
 CC Hummelreich R., Hilbert H., Flagens H., Plöckl E., Li B.-C.,
 CC Herrmann R.;
 CC "Complete sequence analysis of the genome of the bacterium Mycoplasma
 CC pneumoniae."
 CC RT pneumonae."
 CC RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: BINDS SPECIFICALLY TO THE SSRA RNA (TMRNA) AND IS
 CC REQUIRED FOR STABLE ASSOCIATION OF SSRA WITH RIBOSOMES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SSRP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000010; AAB95729.1; -
 DR InterPro: IPR000037; Smpb.
 DR Pfam: PF01668; Smpb; 1.
 DR PRODOM: PD004488; Smpb; 1.
 DR PROSITE: PS01317; SSRP; 1.
 KW RNA-binding; Complete proteome.
 SQ SEQUENCE 147 AA: 17087 MW: 6F4F3E5A4412C40 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 KRVK 6
 DB 74 KRVK 78

RESULT 9
 UBC6_DROME STANDARD; PRT; 151 AA.
 ID UBC6_DROME
 AC P25153; Q9YV70;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE UBIQUITIN-CONJUGATING ENZYME E2-17 KHA (EC 6.3.2.19)
 DE (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN).
 GN UBC6 OR DHR6 OR CG2013.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;

[1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-91219466; PubMed-1902572;
 RA Koken M.H.M., Reynolds P., Bootsma D., Hoeijmakers J.H.J., Prakash S.,
 RT Prakash L.;
 "Dhr6, a Drosophila homolog of the yeast DNA-repair gene RAD6.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3832-3836(1991).
 RP
 RC SEQUENCE FROM N.A.
 RX STRAIN-BERKELEY;
 MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Ceolinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer J.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abil J.F., Abdayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kensington J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Melnikov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klimas I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitslks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang X., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO
 OTHER PROTEINS. REQUIRED FOR POSTREPLICATION REPAIR OF UV-DAMAGED
 DNA.
 CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE - AMP +
 PYROPHOSPHATE + PROTEIN N-UBIQUITYLlysine.
 CC -1- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR
 UBIQUITIN-THIOLESTER FORMATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
 CC STRONGEST, TO YEAST RAD6.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M63792; AAA28308.1; -
 DR EMBL: M64435; AAA28309.1; -
 DR EMBL: M63791; AAA28309.1; JOINED.
 DR EMBL: AE003604; AAF52079.1; -

DR PIR: A39392; A39392.
 DR HSSP: P25865; 2AAR.
 DR Flybase: FBgn0004436; Ubcd6.
 DR InterPro: IPR000608; UBC_conjugat.
 DR Pfam: PF00179; UBC_conj_1.
 DR SMART: SM00212; UBCc_1.
 DR PROSITE: PS00183; UBIQUITIN_CONJUGAT_1; 1.
 DR PROSITE: PS00127; UBIQUITIN_CONJUGAT_2; 1.
 KW Ubiquitin conjugation; ligase; DNA repair; Nuclear protein;
 KM Multigene family.
 FT BINDING 88 UBIQUITIN (BY SIMILARITY).
 FT CONFLICT 99 T -> R (IN REF. 1).
 SQ SEQUENCE 151 AA; 17152 MW; CC4B3592E4A9220 CRC64;
 QY 1 EKKVK 5
 Db 138 EKKVK 142
 RESULT 10
 Y723_METJA STANDARD: PRT; 156 AA.
 ID Y723_METJA
 AC 058133;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR PROTEIN M0723.
 GN M0723.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Reisch C.I.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Ullrich T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
 "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: BELONGS TO THE ASNC FAMILY OF TRANSCRIPTIONAL
 REGULATORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U67519; AAB98719.1; -
 DR TIGR: M0723; -
 DR InterPro: IPR000485; ASNC_trans_reg.
 DR Pfam: PF01037; ASNC_trans_reg; 1.
 DR PRINTS: PRO0033; HTHASNC.
 DR SMART: SM00344; HTH_ASNC; 1.
 DR PROSITE: PS00519; HTH_ASNC_FAMILY; 1.
 KW Hypothetical protein; Transcription regulation; DNA-binding;
 FT Complete proteome.
 KW DNA_BIND 36 55 H-T-H MOTIF (POTENTIAL).

SQ SEQUENCE 156 AA; 17730 MW; F358D2EB692EB480 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 156;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6
11111
Db 52 KRVKK 56

RESULT 11

YR27_CAEEL STANDARD; PRT; 162 AA.
ID YR27_CAEEL
AC Q09343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHEICAL 19.2 KDA PROTEIN W02B3.7 IN CHROMOSOME III.
GN W02B3.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Connell M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U22833; AAA64326.1; -
DR Wormpep; W02B3.7; CE02079.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 10 30 POTENTIAL.
SQ SEQUENCE 162 AA; 19156 MW; 44E1550218ED3611 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 162;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRVKK 5
11111
Db 73 KRVKK 77

RESULT 12
YF81_AQUAE STANDARD; PRT; 167 AA.
ID YF81_AQUAE
AC O67521;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHEICAL PROTEIN AQ_1581.
GN AQ_1581.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Kellar M., Anjaj M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus".
RL Nature 392:353-358(1998).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE000746; AAC07489.1; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 5 27 POTENTIAL.
SQ SEQUENCE 167 AA; 20011 MW; AAFACF36DA64554 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 167;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6
11111
Db 129 KRVKK 133

RESULT 13

HSCB_ECOLI STANDARD; PRT; 171 AA.
ID HSCB_ECOLI
AC P36540;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN HSCB (HSC20).
GN HSCB OR B2527 OR Z3794 OR ECS3393.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=94131939; PubMed=8300516;
RA Kawula T.H., Lelievelt M.J.;
RT "Mutations in a gene encoding a new Hsp70 suppress rapid DNA
RT inversion and bgl activation, but not prov derepression, in hns-1
RT mutant Escherichia coli.";
RL J. Bacteriol. 176:610-619(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Saito G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and

```

RT analysis of its sequence features."
RN DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-0157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grodeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Podaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RN "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RN Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-0157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuwara S., Shiba T., Hattori M., Shitagawa H.;
RN "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RT DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 119-171 FROM N.A.
RC STRAIN-B;
RX MEDLINE-94181532; PubMed-8134349;
RA Seaton B.L., Vickery L.E.;
RN "A gene encoding a DnaK/hsp70 homolog in Escherichia coli.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:2066-2070(1994).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE-97289991; PubMed-9144776;
RA Vickery L.E., Silberg J.J., Ta D.T.;
RN "Hsc66 and Hsc70, a new heat shock cognate molecular chaperone system
RT from Escherichia coli.";
RT Protein Sci. 6:1047-1056(1997).
RN [8]
RP CRYSTALLIZATION.
RX MEDLINE-97445617; PubMed-9300502;
RA Cupp-Vickery J.R., Vickery L.E.;
RN "Crystallization and preliminary X-ray crystallographic properties of
RT Hsc70, a J-motif co-chaperone protein from Escherichia coli.";
RT Protein Sci. 6:2028-2030(1997).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RA MEDLINE-20574017; PubMed-1124030;
RA Cupp-Vickery J.R., Vickery L.E.;
RN "Crystal structure of Hsc70, a J-type co-chaperone from Escherichia
RT coli.";
RT J. Mol. Biol. 304:835-845(2000).
RN [10]
RP FUNCTION: CO-CHAPERONE THAT INTERACTS WITH HSCA AND STIMULATES
RN ITS APASE ACTIVITY.
RN [11]
RP SIMILARITY: CONTAINS 1 J DOMAIN.
RN [12]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation-
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [13]
RN EMBL: U01827; AAA18299.1; -
RN EMBL: AE000339; AAC75580.1; -
RN EMBL: D90883; BAA16421.1; -
RN EMBL: AE005482; AAG57641.1; -
RN EMBL: AP002562; BAB36816.1; -
RN EMBL: U05338; -, NOT_ANNOTATED_CDS.
RN PDB: IFPO; 08-DEC-00.
RN Ecogene; Egi2131; hscB.

```

```

DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; FALSE_NEG.
DR PROSITE: PS50076; DnaJ_2; 1.
RW Chaperone; 3D-structure; Complete proteome.
FT DOMAIN 2 60 J-DOMAIN.
SQ SEQUENCE 171 AA; 20138 MW; 095193FE98AA60C9 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 171;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRYKK 6
DB 119 KRYKK 123

RESULT 14
RM06_ACACA STANDARD; PRT; 181 AA.
ID RM06_ACACA
AC P46765;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L6.
GN RPL6.
OS Acanthamoeba castellanii (Amoeba).
OG Mitochondrion.
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxId=5755;
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 30010 / NEFF;
RX MEDLINE-95147275; PubMed-7844823;
RA Burger G., Plante I., Loneragan K.M., Gray M.W.;
RN "The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba
RT castellanii: complete sequence, gene content and genome
RT organization.";
RT J. Mol. Biol. 245:522-537(1995).
RN [12]
RP SUBCELLULAR LOCATION: MITOCHONDRIAL.
RN [13]
RP SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
RN [14]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation-
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [15]
RN EMBL: U12386; AAD11847.1; -
RN InterPro: IPR000702; Ribosomal_L6.
RN InterPro: IPR002358; Ribosomal_L6_1.
RN Pfam: PF00347; Ribosomal_L6; 1.
RN ProDom: PD002236; Ribosomal_L6; 1.
RN PROSITE: PS00525; RIBOSOMAL_L6_1; 1.
RW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 181 AA; 21772 MW; BB392EF9CFD6B347 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 181;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRYKK 6
DB 148 KRYKK 152

RESULT 15
IF3C_CYACA

```

```

ID IF3C-CYACA STANDARD; PRT; 185 AA.
AC Q9TLX8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSLATION INITIATION FACTOR IF-3, CHLOROPLAST.
GN INFC.
OS Cyanidium caldarium.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium
OX NCBI_TaxID=2771;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RK-1;
RX MEDLINE=20496959; PubMed=11040290;
RA Gloeckner G., Rosenthal A., Valentin K.;
RT "The structure and gene repertoire of an ancient red algal plastid
genome."
RL J. Mol. Evol. 51:382-390(2000).
CC -1- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE
CC EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN
CC FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S
CC SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO THE IF-3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF022186; AAF12954.1; -.
DR HSSP; P03000; 1FTG.
DR InterPro; IPR001288; IF3.
DR Pfam; PF00707; IF3; 1.
DR ProDom; PD002880; IF3; 1.
DR PROSITE; PS00938; IF3; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; Chloroplast.
SQ SEQUENCE 185 AA; 21098 MW; C9732BAA53A8A36F CRC64;

```

```

Query Match 83.3%; Score 5; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. NO. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EKRVK 5
   |||||
Db 82 EKRVK 86

```

```

Search completed: February 26, 2002, 01:53:54
Job time: 423 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:29:56 ; Search time 79.46 Seconds
(without alignments)
5.593 Million cell updates/sec

Title: US-09-819-136-2_COPY_117_122

Perfect score: 6

Sequence: 1 RDRCK 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size: 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	100.0	216	22	Human membrane or Bufo marinus lipoc
2	83.3	183	22	Peptide #3961 enco
3	83.3	304	22	Peptide #4086 enco
4	83.3	304	22	Peptide #3881 enco
5	83.3	304	22	Human EPR-1. Homo
6	83.3	337	16	TRIP-33. Syntheti
7	66.7	9	14	Antipholipid a
8	66.7	9	14	Antibody ACA-6501
9	66.7	9	19	Antibody ACA-6501
10	66.7	9	19	Antibody ACA-6501
11	66.7	9	19	Antibody ACA-6501

12	4	66.7	10	15	AAV37979	Human cERR2 oncoge
13	4	66.7	10	20	AAV45549	Immunogenic peptid
14	4	66.7	12	18	AAW10420	Antipholipid a
15	4	66.7	12	19	AAW45085	Antipholipid a
16	4	66.7	12	19	AAW67248	Antipholipid a
17	4	66.7	13	19	AAW46086	Antipholipid a
18	4	66.7	14	22	AAJ03079	Hepatitis C virus
19	4	66.7	14	22	AAW67040	Human insulin-like
20	4	66.7	15	12	AAAR1328	Grb alpha peptide
21	4	66.7	15	21	AAV98941	HLA class II bindi
22	4	66.7	15	22	AAW88708	HER2/neu DR 3a mot
23	4	66.7	15	22	AAW89075	HER2/neu DR3 bindi
24	4	66.7	15	14	AAAR37462	c-erbB2 hydrophobi
25	4	66.7	16	21	AAW26525	Human IgE C epsilon
26	4	66.7	18	22	AAW62865	Immunogenic peptid
27	4	66.7	24	22	AAW68663	HER-2 B cell epit
28	4	66.7	26	22	AAW76928	Human colon cancer
29	4	66.7	34	21	AAW88023	Human growth hormo
30	4	66.7	34	22	AAE00322	Human cell cycle c
31	4	66.7	41	21	AAW06074	Caenorhabditis ele
32	4	66.7	41	21	AAW06186	Caenorhabditis ele
33	4	66.7	47	22	AAW73675	Human colon cancer
34	4	66.7	58	21	AAW55410	Arabidopsis thailia
35	4	66.7	58	21	AAW61665	Arabidopsis thailia
36	4	66.7	59	22	AAW31218	Peptide #525 enco
37	4	66.7	62	20	AAV13216	Human secreted pro
38	4	66.7	63	21	AAV15192	N-terminal end of
39	4	66.7	64	20	AAV00320	Human secreted pro
40	4	66.7	70	22	AAW60783	Scorpion leluotox
41	4	66.7	71	21	AAW01907	Human secreted pro
42	4	66.7	74	21	AAW41011	zee mays protein f
43	4	66.7	80	21	AAW34617	Peptide #654 enco
44	4	66.7	84	9	AAW80888	Copea trypsin inh
45	4	66.7	91	13	AAW25187	2184 gene clone pr

ALIGNMENTS

RESULT 1	AAW88333	standard; Protein; 216 AA.
ID	AAW88333	
AC	AAW88333	
DT	23-MAY-2001	(first entry)
DE	Human membrane or secretory protein clone PSEC0040.	
KW	Human; secretory protein; membrane protein; vaccine; gene therapy; Rheumatoid arthritis; diabetes.	
OS	Homo sapiens.	
PN	EP1067182-A2.	
PD	10-JAN-2001.	
PF	07-JUL-2000; 2000EP-0114090.	
PR	08-JUL-1999; 99JP-0194179.	
PR	11-JAN-2000; 2000JP-0118775.	
PR	02-MAY-2000; 2000JP-0183766.	
PA	(HELI-) HELIX RES INST.	
PI	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;	
DR	WPT: 2001-093989/11.	
DR	N-PSDB: AAF93760.	
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -	

XX Claim 1; SEQ ID 34; 609pp + CD ROM; English.
 PS
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAF88317 - AAF88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 SQ Sequence 216 AA;
 XX
 XX
 Query Match 100.0%; Score 6; DB 22; Length 216;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RDRCCK 6
 |||||
 Db 117 rdrcck 122

RESULT 2
 AAB90659 standard; Protein; 183 AA.
 ID AAB90659
 XX
 AC AAB90659;
 XX
 DT 01-JUN-2001 (first entry)
 XX
 DE Bufo marinus lipocalin protein, SEQ ID NO: 202.
 XX
 KW Toid; lipocalin; secreted protein; immunomodulatory; antisclerotic;
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
 KW nootropic; anticonvulsant; antialzheimer's; antiparkinsonian;
 KW antimicrobial; vulnerary; vaccine; gene therapy; cancer;
 KW protein coordinate data; infection.
 XX
 OS Bufo marinus.
 XX
 PN WO200121658-A1.
 XX
 PD 29-MAR-2001.
 XX
 PF 22-SEP-2000; 2000WO-US26013.
 XX
 PR 24-SEP-1999; 99US-0155709.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Baker KP, Birse CE, Ebnner R, Piscella M, Komatsoulis GA;
 PI Lafleur DW, Moore RA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
 PI Young PE, Wei P, Florence KA;
 XX

DR WPI; 2001-235311/24.
 XX
 PT Nucleic acids encoding 32 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX
 PS Disclosure; Fig 2; 890pp; English.
 XX
 CC The present sequence is provided in a specification relating to nucleic
 CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic
 CC acid molecules and polypeptides may be used in the prevention, diagnosis
 CC and treatment of diseases such as immune disorders (e.g. multiple
 CC sclerosis), systemic lupus erythematosus and human immuno-deficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. scintar syndrome,
 CC Chagas' cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic
 CC acid molecules may be used to produce the secreted polypeptides. They may
 CC also be used as DNA probes in diagnostic assays to detect and quantitate
 CC the presence of similar nucleic acid sequences in samples. The
 CC polypeptides may be used as antigens in the production of antibodies and
 CC in assays to identify modulators of their expression and activity.
 CC
 SQ Sequence 183 AA;
 XX
 XX
 Query Match 83.3%; Score 5; DB 22; Length 183;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DRCEK 6
 |||||
 Db 81 drcek 85

RESULT 3
 AAM17527 standard; Protein; 304 AA.
 ID AAM17527
 XX
 AC AAM17527;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #3961 encoded by probe for measuring cervical gene expression.
 XX
 DE Peptide #3961 encoded by probe for measuring cervical gene expression.
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 KW Homo sapiens.
 OS
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 XX
 PR 30-JUN-2000; 2000US-0608408.
 XX
 PR 03-AUG-2000; 2000US-0632366.
 XX
 PR 21-SEP-2000; 2000US-0234687.
 XX
 PR 27-SEP-2000; 2000US-0236359.
 XX
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 27; SEQ ID No 22353; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see A110068-A118459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
SQ Sequence 304 AA;

Query Match 83.3%; Score 5; DB 22; Length 304;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRCE 5
| | | | |
Db 197 rdrc 201

RESULT 4
AAM30049
ID AAM30049 standard; Protein; 304 AA.
XX
XX AAM30049;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #4086 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200157272-A2.
PN
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00063.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI: 2001-488897/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 30318; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see A113115-A1157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX
SQ Sequence 304 AA;

Query Match 83.3%; Score 5; DB 22; Length 304;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRCE 5
| | | | |
Db 197 rdrc 201

RESULT 5
AAM05199
ID AAM05199 standard; Protein; 304 AA.
XX
XX AAM05199;
XX
DT 09-OCT-2001 (first entry)
XX
DE Peptide #3881 encoded by probe for measuring breast gene expression.
XX
KW Probe; human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX WO200157270-A2.
PN
XX 09-AUG-2001.
PD
XX
PF 29-JAN-2001; 2001WO-US00061.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI: 2001-476286/51.
DR
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 27; SEQ ID No 13939; 322bp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes
CC (see A1100010-A110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridizes at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosis
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 304 AA;

Query Match 83.3%; Score 5; DB 22; Length 304;

Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDCE 5
11111
Db 197 rdce 201

RESULT 6
AAR6744

ID AAR6744 standard; Protein; 337 AA.

AC AAR6744;

DT 07-DEC-1995 (first entry)

XX Human EPR-1.

DE Effector cell protease receptor-1; EPR-1; lymphocyte;
KW Lymphoproliferative disorder.

OS Homo sapiens.

XX Key Location/Qualifiers
FH Modified-site 2

FT /label= Glycosylation

FT /note= "O-linked glycosylation site"

FT 48..72

FT /note= "Mab 2E1 immunoreactive epitope"

FT 76

FT /label= Glycosylation

FT /note= "O-linked glycosylation site"

FT 125..150

FT /note= "surface-exposed region"

FT 151..176

FT /note= "hydrophobic membrane-spanning domain"

FT 168

FT /label= Glycosylation

FT /note= "N-linked glycosylation site"

FT 170

FT /label= Glycosylation

FT /note= "O-linked glycosylation site"

FT 174

FT /label= Glycosylation

FT /note= "O-linked glycosylation site"

FT 204

FT /note= "chondroitin sulfate attachment site"

FT 265

FT /label= Glycosylation

FT /note= "N-linked glycosylation site"

FT 257..337

FT /note= "C-terminal cytoplasmic tail"

FT Region

XX WO9520655-A.

XX 03-AUG-1995.

XX 18-JAN-1995; 95WO-US00666.

XX 28-JAN-1994; 94US-0189309.

XX (SCRI) SCRIPPS RES INST.

XX Altieri DC;

XX WPI; 1995-275440/36.

XX N-PSDB; AA093052.

XX New effector cell protease receptor-1 protein - used to develop prods.

XX for inducing lymphocyte proliferation or diagnosis and treatment of

XX lympho-proliferative disorders

XX Claim 1; Page 154-155; 181pp; English.

XX Immunoscreening of a human lymphocyte expression library in lambda-
CC gill using Mab 2E1 raised against human EPR-1 yielded clone
CC lambda-104. This was used as a probe in further screening of human
CC cDNA libraries, and a consensus EPR-1 cDNA sequence was deduced from
CC 28 isolated clones. The DNA can be used for prodn. of recombinant
CC EPR-1.
XX
SQ Sequence 337 AA;

Query Match

Best Local Similarity 83.3%; Score 5; DB 16; Length 337;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDCE 5
11111
Db 58 rdce 62

RESULT 7

ID AAR41614 standard; Protein; 9 AA.

AC AAR41614;

DT 25-MAR-1994 (first entry)

XX TRIP-33.

XX Erabutoxin; thrombin; receptor; inhibitor; peptide; TRIP; septicemia;

XX fusion protein; plasmid; pIII; phagemid vector; pT219; syngene;

XX fibrinogen; coagulation; platelet activation; reperfusion damage;

XX neurodegenerative disease; bone resorption; osteoporosis; thrombosis;

XX inflammation; adult respiratory distress syndrome; septic shock.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 4 /note= "D-form residue"

FT WO9318141-A.

FT 16-SEP-1993.

FT 02-MAR-1993; 93WO-US01901.

FT 02-MAR-1992; 92US-0847561.

FT (BIOJ) BIOGEN INC.

XX Freilinger AL; Maraganore JM;

XX WPI; 1993-303456/38.

XX New thrombin receptor antagonist peptide(s) and fusion proteins -

XX used to inhibit thrombin-induced platelet aggregation and smooth

XX muscle cell proliferation

XX Example 38; Page 46; 95pp; English.

XX The sequences given in AAR41583-633 are thrombin receptor inhibitor

XX peptides (TRIPs). TRIPs competitively bind to and inhibit the

XX thrombin receptor without affecting thrombin activities toward

XX fibrinogen and other proteins involved in coagulation. They can

XX be used for inhibiting thrombin-induced platelet activation and

XX inhibiting thrombin-induced smooth muscle cell proliferation. They

XX can be used for treating and preventing neurodegenerative disease,

XX treating and preventing bone resorption (eg. in osteoporosis),

XX treating and preventing thrombin-induced inflammation, eg. in adult

XX respiratory distress syndrome, septic shock, septicemia and

XX reperfusion damage) and for inhibiting platelet-dependent thrombosis.

CC Fusion proteins containing these sequences may also act as thrombin
CC receptor antagonists.

SO Sequence 9 AA:

Query Match 66.7%; Score 4; DB 14; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
Db 3 rcek 6

RESULT 8

AA041625 standard; Peptide: 9 AA.

AA041625;

25-MAR-1994 (first entry)

TRIP-44.

Erythrocyte; thrombin; receptor; inhibitor; peptide; TRIP; septicemia;
fusion protein; plasma; fibrin; phagocytosis; syngene;
fibrinogen; coagulation; platelet activation; reperfusion damage;
neurodegenerative disease; bone resorption; osteoporosis; thrombosis;
inflammation; adult respiratory distress syndrome; septic shock.

Synthetic.

Key Location/Qualifiers

Misc-difference 4 /note= "D-form residue"

Misc-difference 5 /note= "D-form residue"

W09318141-A.

16-SEP-1993.

02-MAR-1993; 93WO-US01901.

02-MAR-1992; 92US-0847561.

(BIOJ) BIOGEN INC.

Freilinger AL, Maraganore JM;

WPI: 1993-303456/38.

New thrombin receptor antagonist peptide(s) and fusion proteins -
used to inhibit thrombin-induced platelet aggregation and smooth
muscle cell proliferation

Example 49; Page 48; 95pp; English.

The sequences given in AA041583-633 are thrombin receptor inhibitor
peptides (TRIPs). TRIPs competitively bind to and inhibit the
thrombin receptor without affecting thrombin activities toward
fibrinogen and other proteins involved in coagulation. They can
be used for inhibiting thrombin-induced platelet activation and
inhibiting thrombin-induced smooth muscle cell proliferation. They
can be used for treating and preventing neurodegenerative disease,
treating and preventing bone resorption (eg. in osteoporosis),
respiratory distress syndrome, septic shock, septicemia and
fusion proteins containing these sequences may also act as thrombin
receptor antagonists.

SO Sequence 9 AA:

Query Match 66.7%; Score 4; DB 14; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
Db 3 rcek 6

RESULT 9

AA010403 standard; peptide: 9 AA.

AA010403;

21-AUG-1997 (first entry)

Antiphospholipid analogue peptide ACA-6501 phage library clone 2H4.

APL; antibody mediated pathology; systemic lupus erythematosus;

antiphospholipid antibody syndrome; stroke; foetal loss; thrombosis.

Synthetic.

W09640197-A1.

19-DEC-1996.

06-JUN-1996; 96WO-US09976.

07-JUN-1995; 95US-0482651.

(LJOL-) LA JOLLA PHARM CO.

Jones DS, Marquis DM, Victoria EJ, Yu L;

WPI: 1997-051886/05.

New anti-phospholipid analogues - used to develop prods. for the
diagnosis and treatment of apl antibody-mediated pathologies such as
systemic lupus erythematosus

Disclosure: Page 27; 118pp; English.

The present sequence is an antiphospholipid (APL) analogue clone
2H4 screened from ACA-6501 phage library. Peptides are derived
from screening random peptide libraries to identify
the sequences which best mimic the epitopes recognised by APL
antibodies. A method has been produced for identifying analogues of
key epitopes recognised by APL antibodies in patients suffering from
CC PAPS, APS and other APL antibody-mediated diseases. The products and
CC methods can be used for the diagnosis and treatment of APL antibody-
mediated pathologies such as systemic lupus erythematosus,
antiphospholipid antibody syndrome, recurrent stroke, recurrent foetal
loss or thrombosis.

Sequence 9 AA:

Query Match 66.7%; Score 4; DB 18; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRC 4
Db 6 rdrc 9

RESULT 10

AA067126

ID AAW67126 standard; peptide; 9 AA.
 XX AAW67126;
 AC
 XX
 DT 18-DEC-1998 (first entry)
 XX
 DE Antibody ACA-6501 epitope isolated from a phage library screen.
 XX
 KW antiphospholipid; aPL; analogue; stroke; foetal loss; B cell; epitope;
 KW thrombosis; antibody ACA-6501; conjugate; phage library screen;
 KW biopanning; micropanning.
 XX
 OS Synthetic.
 XX
 PN W09746251-A1.
 XX
 PD 11-DEC-1997.
 XX
 PF 06-JUN-1997; 97WO-US10075.
 XX
 PR 05-DEC-1996; 96US-0760508.
 PR 06-JUN-1996; 96US-0660092.
 XX
 PA (LJOL-) LA JOLLA PHARM CO.
 XX
 PI Jones DS, Marguis DM, Victoria EJ, Yu L;
 XX
 DR WPI; 1998-051902/05.
 XX
 PT New anti:phospholipid analogues and conjugates containing them - are
 PT useful in treatment of disorders mediated by antiphospholipid
 PT antibody, such as stroke or foetal loss
 XX
 PS Disclosure; Page 26; 155pp; English.
 XX
 CC The invention relates to antiphospholipid (aPL) analogues that bind
 CC specifically to B cells to which an aPL epitope binds. Optimised
 CC analogues lack T cell epitope(s) and are useful as conjugates for
 CC treating aPL antibody-mediated diseases. Conjugates comprising aPL
 CC analogues and non-immunogenic valency platform molecules are provided as
 CC are novel non-immunogenic valency platform molecules and linkers.
 CC Methods of preparing and identifying the analogues, methods of treatment
 CC and diagnostic immunosays for preparing conjugates of the analogues
 CC and conjugates are useful for treatment of aPL antibody-mediated diseases,
 CC such as stroke, foetal loss, antiphospholipid antibody syndrome or
 CC thrombosis. Antibody ACA-6501, from a patient with a history of recurrent
 CC stroke, foetal loss, lupus and aortic valve replacements, was used in
 CC phage library screens. Sequences AAW67097-W67132 represent epitopes that
 CC were obtained in phage that micropanned.
 XX
 SQ Sequence 9 AA;
 XX
 Query Match 66.7%; Score 4; DB 19; Length 9;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 RDRC 4
 DB 6 rdrc 9
 XX
 RESULT 11
 ID AAW67133 standard; peptide; 9 AA.
 XX
 AC AAW67133;
 XX
 DT 18-DEC-1998 (first entry)
 XX
 DE Antibody ACA-6501 epitope isolated from a phage library screen.
 XX

KW antiphospholipid; aPL; analogue; stroke; foetal loss; B cell; epitope;
 KW thrombosis; antibody ACA-6501; conjugate; phage library screen;
 KW biopanning; micropanning.
 XX
 OS Synthetic.
 XX
 PN W09746251-A1.
 XX
 PD 11-DEC-1997.
 XX
 PF 06-JUN-1997; 97WO-US10075.
 XX
 PR 05-DEC-1996; 96US-0760508.
 PR 06-JUN-1996; 96US-0660092.
 XX
 PA (LJOL-) LA JOLLA PHARM CO.
 XX
 PI Jones DS, Marguis DM, Victoria EJ, Yu L;
 XX
 DR WPI; 1998-051902/05.
 XX
 PT New anti:phospholipid analogues and conjugates containing them - are
 PT useful in treatment of disorders mediated by antiphospholipid
 PT antibody, such as stroke or foetal loss
 XX
 PS Disclosure; Page 26; 155pp; English.
 XX
 CC The invention relates to antiphospholipid (aPL) analogues that bind
 CC specifically to B cells to which an aPL epitope binds. Optimised
 CC analogues lack T cell epitope(s) and are useful as conjugates for
 CC treating aPL antibody-mediated diseases. Conjugates comprising aPL
 CC analogues and non-immunogenic valency platform molecules are provided as
 CC are novel non-immunogenic valency platform molecules and linkers.
 CC Methods of preparing and identifying the analogues, methods of treatment
 CC and diagnostic immunosays for preparing conjugates of the analogues
 CC and conjugates are useful for treatment of aPL antibody-mediated diseases,
 CC such as stroke, foetal loss, antiphospholipid antibody syndrome or
 CC thrombosis. Antibody ACA-6501, from a patient with a history of recurrent
 CC stroke, foetal loss, lupus and aortic valve replacements, was used in
 CC phage library screens. Sequence AAW67133 represents an epitope that
 CC was isolated from a phage library screen using antibody ACA-6501.
 XX
 SQ Sequence 9 AA;
 XX
 Query Match 66.7%; Score 4; DB 19; Length 9;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 RDRC 4
 DB 6 rdrc 9
 XX
 RESULT 12
 ID AAY37979 standard; peptide; 10 AA.
 XX
 AC AAY37979;
 XX
 DT 29-SEP-1999 (first entry)
 XX
 DE Human CEBB2 oncogene-derived HLA-binding peptide.
 XX
 KW Immunogen; HLA; human leukocyte antigen; binding motif; antiviral;
 KW MHC; major histocompatibility complex; viral infection; anticancer;
 KW prostate cancer; lymphoma; hepatitis; AIDS; diagnostic; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN W09403205-A1.
 XX

PD 17-FEB-1994.
 XX
 PF 06-AUG-1993; 93WO-US07421.
 XX
 PR 05-MAR-1993; 93US-0027746.
 PR 07-AUG-1992; 92US-0926666.
 XX
 PA (CYTE-) CYTEL CORP.
 XX
 PI Cells E, Grey HM, Kubo RT, Sette A;
 XX WPI; 1994-065403/08.
 DR
 XX Peptide which specifically binds selected MHC allele - used to
 PT induce an immune response for treatment or prevention of viral
 PT infection or cancer, or for diagnosis
 PS
 XX Disclosure; Page 103; 150pp; English.
 XX
 XX The sequence is a specific example of a group of new immunogenic
 CC peptides having an HLA-A3.2, HLA-A1, HLA-A11 or HLA-A24.1 binding
 CC motif. For example, the peptides having an HLA-A3.2 binding motif
 CC each have 9-10 residues and contain, from the N-terminus to the
 CC C-terminus, (a) a first conserved residue selected from L, M, I,
 CC V, S, A, T, F, C, G, D and E and (b) a second conserved residue of
 CC K, R, Y, H or F, where the first and second conserved residues are
 CC separated by 6-7 residues. The peptides are capable of binding
 CC selected MHC molecules and inducing an immune response. They can be
 CC used to treat and/or prevent viral infection and cancer, e.g. prostate
 CC cancer, lymphoma, hepatitis or AIDS. They can also be used to produce
 CC antibodies for use as diagnostic or therapeutic agents. The peptides
 CC can also be used as diagnostic agents.
 XX
 SQ Sequence 10 AA:

Query Match 66.7%; Score 4; DB 15; Length 10;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RCEK 6
 ||||
 Db 4 rcek 7

RESULT 13
 AAY45549
 ID AAY45549 standard; Peptide; 10 AA.
 XX
 AC AAY45549;
 XX
 DT 01-DEC-1999 (first entry)
 XX
 DE Immunogenic peptide having a human leukocyte antigen binding motif #160.
 XX
 KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
 KW immune response; T cell activation; major histocompatibility complex;
 KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
 KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
 KW vaccine; immunisation.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN MO9945954-A1.
 XX
 PD 16-SEP-1999.
 XX
 PF 13-MAR-1998; 98WO-US05039.
 XX
 PR 13-MAR-1998; 98WO-US05039.
 XX
 PA (EPIM-) EPIMUNE INC.

XX Sette A, Kubo RT, Sidney J, Cells E, Grey HM, Southwood S;
 PI WPI; 1999-551214/46.
 XX
 DR
 XX New immunogenic peptides with HLA binding motif, useful in treatment
 PT and diagnosis of cancers and viral diseases -
 PT
 XX
 PS Claim 1; Page 33; 150pp; English.
 XX
 XX AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
 CC having a human major histocompatibility complex (MHC) Class I (also
 CC known as human leukocyte antigen (HLA)) binding motif. The immunogenic
 CC peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
 CC HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
 CC response against the antigen from which the peptide is derived.
 CC Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
 CC normally induced by an antigen in the form of a peptide fragment bound
 CC to a HLA molecule, rather than the intact foreign antigen itself, and
 CC are particularly important in tumour rejection and in fighting viral
 CC infections. The peptides are therefore useful therapeutically to treat
 CC or prevent viral infections and cancers in mammals (especially humans)
 CC e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
 CC They can be administered as vaccines to elicit an immune response in
 CC individuals susceptible or otherwise at risk of viral infection or
 CC cancer, or used to treat chronic or acute conditions. They are also
 CC useful diagnostically, and can be used to induce a cytotoxic T cell
 CC response, by contacting a cytotoxic T cell with the peptide e.g. to
 CC produce CTLs ex vivo for infusion back into a patient. The
 CC polynucleotides encoding the immunogenic peptides are also useful
 CC therapeutically and for immunisation as above.
 XX
 SQ Sequence 10 AA:

Query Match 66.7%; Score 4; DB 20; Length 10;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RCEK 6
 ||||
 Db 4 rcek 7

RESULT 14
 AAW10420
 ID AAW10420 standard; peptide; 12 AA.
 XX
 AC AAW10420;
 XX
 DT 21-AUG-1997 (first entry)
 XX
 DE Antiphospholipid analogue cyclic peptide CB2.
 XX
 KW aPL; antibody mediated pathology; systemic lupus erythematosus;
 KW antiphospholipid antibody syndrome; stroke; foetal loss; thrombosis.
 KW
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Disulfide-bond 3..11
 FT /note- "Cyclised via disulphide bond"
 XX
 PN WO9640197-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09976.
 XX
 PR 07-JUN-1995; 95US-0482651.
 XX
 PA (LJOL-) LA JOLLA PHARM CO.

PI Jones DS, Marquis DM, Victoria EJ, Yu L;
 XX WPI; 1997-051886/05.
 DR
 XX
 PT New anti-phospholipid analogues - used to develop prods. for the
 PT diagnosis and treatment of aPL antibody-mediated pathologies such as
 PT systemic lupus erythematosus
 XX
 XX Example 18; Page 82; 118pp; English.
 XX
 CC The present sequence is an antiphospholipid (aPL) analogue, cyclised via
 CC a disulphide bond, peptide derived from peptide CB2. Peptides are
 CC derived from screening random peptide libraries to identify the
 CC sequences which best mimic the epitopes recognised by aPL antibodies.
 CC A method has been produced for identifying analogues of key epitopes.
 CC recognised by aPL antibodies in patients suffering from PAPS, APS and
 CC other aPL antibody-mediated diseases. The products and methods can be
 CC used for the diagnosis and treatment of aPL antibody-mediated
 CC pathologies such as systemic lupus erythematosus, antiphospholipid
 CC antibody syndrome, recurrent stroke, recurrent foetal loss or
 CC thrombosis.
 CC
 CC Sequence 12 AA:
 CC
 SQ

Query Match 66.7%; Score 4; DB 18; Length 12;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RDRC 4
 ||||
 Db 8 rdrc 11

RESULT 15
 AAM46085
 ID AAM46085 standard; peptide; 12 AA.
 XX
 AC AAM46085;
 XX
 DT 18-DEC-1998 (first entry)
 XX
 DE Antiphospholipid analogue.
 XX
 KW antiphospholipid; aPL; analogue; stroke; foetal loss; B cell; epitope;
 KW thrombosis; antibody; conjugate.
 XX
 OS Synthetic.
 XX
 PN WO9746251-A1.
 XX
 PD 11-DEC-1997.
 XX
 PF 06-JUN-1997; 97WO-US10075.
 XX
 PR 05-DEC-1996; 96US-0760508.
 PR 06-JUN-1996; 96US-0660092.
 XX
 PA (LJOL-) LA JOLLA PHARM CO.
 XX
 PI Jones DS, Marquis DM, Victoria EJ, Yu L;
 XX
 DR WPI; 1998-051902/05.
 XX
 PT New anti-phospholipid analogues and conjugates containing them - are
 PT useful in treatment of disorders mediated by antiphospholipid
 PT antibody, such as stroke or foetal loss
 XX
 PS Claim 5; 112; 155pp; English.
 CC
 CC The invention relates to antiphospholipid (aPL) analogues that bind
 CC specifically to B cells to which an aPL epitope binds. Optimised
 CC analogues lack T cell epitope(s) and are useful as conjugates for

CC treating aPL antibody-mediated diseases. Conjugates comprising aPL
 CC analogues and non-immunogenic valency platform molecules are provided as
 CC are novel non-immunogenic valency platform molecules and linkers. Methods
 CC of preparing and identifying the analogues, methods of treatment and
 CC methods and compositions for preparing conjugates of the analogues and
 CC diagnostic immunoassays for aPL antibodies are disclosed. The conjugates
 CC are useful for treatment of aPL antibody-mediated diseases, such as
 CC stroke, foetal loss, antiphospholipid antibody syndrome or thrombosis.
 CC The present sequence represents a specifically claimed antiphospholipid
 CC analogue.
 CC
 CC Sequence 12 AA:
 CC
 SQ

Query Match 66.7%; Score 4; DB 19; Length 12;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RDRC 4
 ||||
 Db 8 rdrc 11

Search completed: February 26, 2002, 01:50:46
 Job time: 1250 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:50:48 : Search time 79.46 Seconds

(without alignments)
5.593 Million cell updates/sec

Title: US-09-819-136-2_COPY_50_55

Perfect score: 6

Sequence: 1 ERECSR 6

Scoring table:

Gapop 60.0 , Gapext 60.0

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	100.0	216	AA888333	Human membrane or
2	83.3	10	AAW17407	Chicken anemia v1
3	83.3	52	AA604828	Arabidopsis thalia
4	83.3	83	AA30425	Mature nematode ex
5	83.3	84	AA30423	Mature nematode ex
6	83.3	102	AA30401	Nematode extracted
7	83.3	102	AA15295	A. duodenale nemat
8	83.3	108	AA35635	Trypophan aporepr
9	83.3	112	AA651780	Arabidopsis thalia
10	83.3	115	AA30218	Arabidopsis thalia
11	83.3	150	AA31374	Human bone morphog

12	5	83.3	150	17	AA86903	Human BMP-9 polype
13	5	83.3	150	21	AA82272	Human bone morphog
14	5	83.3	161	20	AA30438	Mature nematode ex
15	5	83.3	162	20	AA30436	Mature nematode ex
16	5	83.3	162	21	AA15321	A. caninum nematod
17	5	83.3	162	21	AA651782	Arabidopsis thalia
18	5	83.3	180	17	AA891713	Adunap7. Ancylost
19	5	83.3	180	17	AA30388	Nematode extracted
20	5	83.3	181	17	AA891711	Acanth45. Ancylos
21	5	83.3	181	20	AA30409	Nematode extracted
22	5	83.3	185	22	AA39104	Human polypeptide
23	5	83.3	195	21	AA651781	Arabidopsis thalia
24	5	83.3	216	16	AA65200	VP2 protein. Chic
25	5	83.3	216	17	AA88498	VP2 of chicken inf
26	5	83.3	244	22	AA40890	Human polypeptide
27	5	83.3	310	22	AA89456	Human protein sequ
28	5	83.3	382	21	AA804113	Phylyl/prelytrins
29	5	83.3	385	20	AA35524	Chlamydia pneumon
30	5	83.3	430	20	AA31745	Mycobacterium tube
31	5	83.3	459	21	AA82081	Chlamydia pneumon
32	5	83.3	556	22	AA893319	Human protein sequ
33	5	83.3	568	21	AA818230	Plasmodium falcipa
34	5	83.3	571	20	AA36845	Protein involved i
35	5	83.3	629	21	AA647486	Arabidopsis thalia
36	5	83.3	639	21	AA647485	Arabidopsis thalia
37	5	83.3	668	21	AA647484	Arabidopsis thalia
38	5	83.3	728	22	AA39515	Human polypeptide
39	5	83.3	728	22	AA41301	Human polypeptide
40	5	83.3	1868	21	AA95442	Caenorhabditis ele
41	5	83.3	2034	20	AA906300	Human activated ca
42	5	83.3	2038	20	AA14595	Human T-type volta
43	5	83.3	2044	20	AA14594	Human T-type volta
44	5	83.3	2353	20	AA06288	Human activated ca
45	5	83.3	2353	20	AA06289	Human activated ca

ALIGNMENTS

RESULT	ID	AA888333	standard; Protein; 216 AA.
AC	AA888333;		
DT	23-MAY-2001	(first entry)	
DE	Human membrane or secretory protein clone PSEC0040.		
KW	Human; secretory protein; membrane protein; vaccine; gene therapy;		
KW	rheumatoid arthritis; diabetes.		
OS	Homo sapiens.		
PN	EP1067182-A2.		
PD	10-JAN-2001.		
PF	07-JUL-2000; 2000EP-0114090.		
PR	08-JUL-1999; 99JP-0194179.		
PR	11-JAN-2000; 2000JP-0118775.		
PR	02-MAY-2000; 2000JP-0183766.		
PA	(HELI-) HELIX RES INST.		
PI	Oca T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;		
DR	WPI; 2001-093989/11.		
DR	N-PSDB; AAF93760.		
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -		

XX Claim 1; SEQ ID 34; 609pp + CD ROM; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAF88317 - AAF88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.

SO Sequence 216 AA;

Query Match

Best Local Similarity 100.0%; Score 6; DB 22; Length 216;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECSR 6
 |||||
 Db 50 erecsr 55

RESULT 2

AAW17407
 ID AAW17407 standard; peptide; 10 AA.

AC AAW17407;

DT 18-JUN-1997 (first entry)

DE Chicken anaemia virus immunogenic peptide #7.

KW Immunogenic polypeptide; CAV; serum; antibody; diagnosis;

KW Chicken anaemia virus; vaccine.

OS Chicken anaemia virus.

PN JP09071599-A.

PD 18-MAR-1997.

PF 06-SEP-1995; 95JP-0229004.

PR 06-SEP-1995; 95JP-0229004.

XX (NISK) NIPPON SEIBUTSU KAGAKU KENKYUSHO ZH.

PA WPI; 1997-231187/21.

PT Polypeptide(s) highly reactive to chick anaemia virus-infected chicken

PT serum - are used for antibody and vaccine production for protection

PS against chicken anaemia virus

PS Claim 1; Page 13; 26pp; English.

CC The sequences given in AAW17401-17 are immunogenic polypeptides which
 CC have a high reactivity to CAV-infected chicken serum. These peptides
 CC and antibodies against them can be used for the diagnosis of chicken
 CC anaemia virus (CAV) infection or in vaccines for preventing CAV
 CC infection. The polypeptides have a high reactivity to CAV-infected
 CC chicken serum.

SO Sequence 10 AA;

Query Match

Best Local Similarity 83.3%; Score 5; DB 18; Length 10;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 |||||
 Db 1 recsr 5

RESULT 3

AAO04828
 ID AAO04828 standard; Protein; 52 AA.

AC AAO04828;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 1002.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151330.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.

PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match

Best Local Similarity 100.0%; Score 5; DB 21; Length 52;
Pred. No. 13;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 12 recsr 16

RESULT 4

AA30425
ID AAY30425 standard; Protein; 83 AA.

AC AAY30425;

DT 15-NOV-1999 (first entry)

DE Mature nematode extracted anticoagulant protein AdunAP7d1.

KW Nematode extracted anticoagulant protein; NAP: anticoagulant;
serine protease inhibitor; NAP domain; factor VIIa/TF.

OS Ancylostoma duodenale.

PN US5955294-A.

PD 21-SEP-1999.

PF 19-APR-1996; 96US-0634641.

PR 19-APR-1996; 96US-0634641.

PR 18-OCT-1994; 94US-0326110.

PR 05-JUN-1995; 95US-0461965.

PR 05-JUN-1995; 95US-0465380.

PR 05-JUN-1995; 95US-0486397.

PR 05-JUN-1995; 95US-0486397.

PR 17-OCT-1995; 95WO-US13231.

PA (CORV-) CORVAS INT INC.

PI Bergum PW, Ganssems YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PBH;
PI Vlaeuk GP;

DR WPI: 1999-539569/45.

PT Screening an isolated protein for Nematode-extracted Anticoagulant

PT Protein domains

PS Disclosure; Columns 137-138; 197pp; English.

The present sequence represents a nematode extracted anticoagulant protein (NAP). The protein has activity as an anticoagulant and/or serine protease inhibitor. The protein contains at least one NAP domain which has selective inhibitory activity for factor VIIa/TF. The specification describes a method for screening an isolated protein at least one domain for factor VIIa/TF selective inhibitory activity. The method comprises determining the time to clotting effected by a concentration of the isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo activated partial thromboplastin time (APTT) assay; calculating the PT and APTT assay, with respect to a baseline clotting value for each assay, where prolongation of clotting is calculated as fold elevation of clotting time relative to a baseline clotting value; and calculating a PT to APTT prolongation ratio, where a ratio at least one is indicative of factor VIIa/TF inhibitory activity. The method is useful for determining if a protein has factor VIIa/TF inhibitory activity.

Sequence 83 AA;

Query Match

Best Local Similarity 100.0%; Score 5; DB 20; Length 83;
Pred. No. 19;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 43 recsr 47

RESULT 5

AA30423
ID AAY30423 standard; Protein; 84 AA.

AC AAY30423;

DT 15-NOV-1999 (first entry)

DE Mature nematode extracted anticoagulant protein AcanAP45d1.

KW Nematode extracted anticoagulant protein; NAP: anticoagulant;
serine protease inhibitor; NAP domain; factor VIIa/TF.

OS Ancylostoma caninum.

PN US5955294-A.

PD 21-SEP-1999.

PF 19-APR-1996; 96US-0634641.

PR 19-APR-1996; 96US-0634641.

PR 18-OCT-1994; 94US-0326110.

PR 05-JUN-1995; 95US-0461965.

PR 05-JUN-1995; 95US-0465380.

PR 05-JUN-1995; 95US-0486397.

PR 17-OCT-1995; 95WO-US13231.

PA (CORV-) CORVAS INT INC.

PI Bergum PW, Ganssems YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PBH;
PI Vlaeuk GP;

DR WPI: 1999-539569/45.

PT Screening an isolated protein for Nematode-extracted Anticoagulant

PT Protein domains

PS Disclosure; Columns 135-136; 197pp; English.

The present sequence represents a nematode extracted anticoagulant protein (NAP). The protein has activity as an anticoagulant and/or serine protease inhibitor. The protein contains at least one NAP domain which has selective inhibitory activity for factor VIIa/TF. The specification describes a method for screening an isolated protein at least one domain for factor VIIa/TF selective inhibitory activity. The method comprises determining the time to clotting effected by a concentration of the isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo activated partial thromboplastin time (APTT) assay; calculating the PT and APTT assay, with respect to a baseline clotting value for each assay, where prolongation of clotting is calculated as fold elevation of clotting time relative to a baseline clotting value; and calculating a PT to APTT prolongation ratio, where a ratio at least one is indicative of factor VIIa/TF inhibitory activity. The method is useful for determining if a protein has factor VIIa/TF inhibitory activity.

SQ Sequence 84 AA;

Query Match 83.3%; Score 5; DB 20; Length 84;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 |||||
 Db 44 recsr 48

RESULT 6

AA30401

ID AAY30401 standard; Protein; 102 AA.

AC AAY30401;

DT 15-NOV-1999 (first entry)

DE Nematode extracted anticoagulant protein AduNAP7.

KW Nematode extracted anticoagulant protein; NAP; anticoagulant;

KM serine protease inhibitor; NAP domain; factor VIIa/TF.

XX Ancylostoma duodenale.

OS US5955294-A.

PN 21-SEP-1999.

PD 19-APR-1996; 96US-0634641.

PF 19-APR-1996; 96US-0634641.

PR 18-OCT-1994; 94US-0326110.

PR 05-JUN-1995; 95US-0461965.

PR 05-JUN-1995; 95US-0465380.

PR 05-JUN-1995; 95US-0486397.

PR 17-OCT-1995; 95WO-US13231.

XX (CORV-) CORVAS INT INC.

PA Bergum PM, Gansemans YGJ, Jespers LS, Laroche YR;

PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PEH;

PI Vlaeuk GP;

XX WPI; 1999-539569/45.

XX Screening an isolated protein for Nematode-extracted Anticoagulant

XX Protein domains

XX Disclosure; Column 111-112; 197pp; English.

XX The present sequence represents a nematode extracted anticoagulant

XX protein (NAP). The protein has activity as an anticoagulant and/or serine

XX protease inhibitor. The protein contains at least one NAP domain which

XX has selective inhibitory activity for factor VIIa/TF. The specification

XX describes a method for screening an isolated protein at least one domain

XX for factor VIIa/TF selective inhibitory activity. The method comprises

XX determining the time to clotting effected by a concentration of the

XX isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo

XX activated partial thromboplastin time (APTT) assay; calculating

XX prolongation of clotting effected by the isolated protein in each of

XX the PT and APTT assay, with respect to a baseline clotting value for

XX each assay, where prolongation of clotting is calculated as fold

XX elevation of clotting time relative to a baseline clotting value, where

XX a doubling of clotting time is deemed a two-fold elevation; and

XX calculating a PT to APTT prolongation ratio, where a ratio at least

XX one is indicative of factor VIIa/TF inhibitory activity. The method is

XX useful for determining if a protein has factor VIIa/TF inhibitory

XX activity.

SQ Sequence 102 AA;

Query Match 83.3%; Score 5; DB 20; Length 102;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 |||||
 Db 62 recsr 66

RESULT 7

AAB15295

ID AAB15295 standard; Protein; 102 AA.

AC AAB15295;

DT 19-DEC-2000 (first entry)

DE A. duodenale nematode-extracted anticoagulant protein AduNAP7.

KW Nematode-extracted anticoagulant protein; AduNAP7; blood clotting;

KM canine hookworm; thrombosis; vaccine.

XX Ancylostoma duodenale.

OS Key Location/Qualifiers

FH Peptide 1..19

FT Protein /note="pre-AduNAP7 peptide"

FT Protein /note="mature AduNAP7"

XX US6087487-A.

XX 11-JUL-2000.

XX 12-FEB-1999; 99US-0249451.

XX 17-OCT-1995; 95WO-US13231.

XX 17-APR-1997; 97US-0809455.

XX 18-OCT-1994; 94US-0326110.

XX 05-JUN-1995; 95US-0461965.

XX 05-JUN-1995; 95US-0465380.

XX 05-JUN-1995; 95US-0486397.

XX 05-JUN-1995; 95US-0486397.

XX (CORV-) CORVAS INT INC.

XX Lauwereys MJ, Stanssens PEH, Jespers LS, Gansemans YGJ, Moyle M;

XX Bergum PM, Messens JHL, Laroche YR, Vlaeuk GP;

XX WPI; 2000-531359/48.

XX N-PSDB; AAY73368.

XX New CDNA molecule encoding a protein having factor Xa inhibitory

XX activity for preventing and treating blood clotting disorders,

XX PT comprises nematode-extracted anticoagulant protein domains -

XX Disclosure; Fig 11; 197pp; English.

XX The present sequence is the Ancylostoma duodenale nematode-extracted

XX anticoagulant protein AduNAP7. Proteins of this kind have been shown to

XX be effective at preventing blood clotting without causing excessive

XX bleeding. The protein can be used in blood collection tubes to aid the

XX isolation of plasma from the blood, to prevent thrombosis which may be

XX linked to the rupture of an atherosclerotic plaque, acute myocardial

XX infarction, angina, thrombolytic therapy, percutaneous transluminal

XX coronary angioplasty, disseminated intravascular coagulopathy, infection,

XX cancer and septic shock, and to produce antibodies. In the latter

XX instance, the antibodies can be raised in order to detect infection by

XX nematodes (the coding sequence can also be used for this) or as

XX diagnostic tests. The protein can also be used as a vaccine against

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160776.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 83.3%; Score 5; DB 21; Length 112;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RECSR 6

Db 30 recsr 34

RESULT 10

AAG30218
ID AAG30218 standard; Protein; 115 AA.
XX
AC AAG30218;
XX

DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 36088.
DE
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hydriolisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123380.
PR 09-MAR-1999; 99US-0123348.
PR 22-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132483.
PR 14-MAY-1999; 99US-0132526.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

```

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      83.3%; Score 5; DB 21; Length 115;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 RECSR 6
   |||||
Db 33 recsr 37

```

```

RESULT 11
AAR31374
ID AAR31374 standard; Protein; 150 AA.
XX
AC AAR31374;
XX
DT 24-JUN-1993 (first entry)
XX
DE Human bone morphogenic protein-9 (BMP-9).
XX
KM bone formation; cartilage formation; wound healing; tissue repair;
KM surgery; fracture treatment; periodontal disease; osteoporosis;
KM increase neuronal survival; transplantation; nerves; nervous system.

```

```

XX Homo sapiens.
OS
XX
PN W09300432-A.
XX
PD 07-JAN-1993.
XX
PF 25-JUN-1992; 92MO-US05374.
XX
PR 25-JUN-1991; 91US-0720590.
XX
XX (GEMV ) GENETICS INST INC.
XX
XX Celeste AJ, Wozney JM;
XX WPI; 1993-036379/04.
XX DR N-PDB; AAQ35244.
XX
PT New BMP-9 polypeptide(s) induce bone and cartilage formation -
PT used for treating osteoporosis and fractures, healing wounds and
PT increasing neuronal survival
XX
XX Example 2; Fig 3; 60pp; English.
XX
CC This sequence is human bone morphogenic protein-9. The protein can
CC be used to induce bone and/or cartilage formation and in wound healing
CC and tissue repair. It can be used in surgery of the treatment of
CC fractures, periodontal disease or osteoporosis. The protein can also
CC increase neuronal survival and can be used in transplantation and
CC treatment of conditions exhibiting a decrease in neuronal survival.
XX
SQ Sequence 150 AA;

```

```

Query Match      83.3%; Score 5; DB 14; Length 150;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 RECSR 6
   |||||
Db 2 recsr 6

```

```

RESULT 12
AAR86903
ID AAR86903 standard; Protein; 150 AA.
XX
AC AAR86903;
XX
DT 10-MAY-1996 (first entry)
XX
DE Human BMP-9 polypeptide.
XX
KM BMP-9; bone; cartilage; wound healing; liver; tissue repair.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 40..41
XX FT Cleavage-site /note="proteolytic cleavage site"
XX FT Protein 41..150
XX FT /label= Mat_protein
XX
XX W09533830-A1.
XX
XX 14-DEC-1995.
XX
XX 05-JUN-1995; 95MO-US07084.
XX
XX 06-JUN-1994; 94US-0254353.
XX
XX (GEMV ) GENETICS INST INC.
XX

```

PI Celeste AJ, Rosen VA, Song JJ, Thies SR, Wozney JM;
XX WPI; 1996-040235/04.
DR N-PSDB; AAT13143.
XX
PT BMP-9 protein able to induce bone and cartilage formation - also
PT useful in wound healing, tissue repair and for inducing hepatic
PT growth and function
XX
PS Claim 1; Page 45-46; 75pp; English.
XX
CC Human BMP-9 polypeptide (AAR6903) is the product of a DNA fragment
CC (AA074084) isolated from a human genomic DNA library. Cleavage of
CC the polypeptide at the consensus proteolytic processing site
CC gives a 110-amino acid mature protein that dimerizes to form a
CC homodimeric, active species. Recombinant BMP-9 can be obt. by
CC expression of the genomic DNA in prokaryotic or eucaryotic hosts.
CC It has therapeutic applns. in the formation of bone and
CC cartilage, wound healing and tissue repair.
XX
SQ Sequence 150 AA;

Query Match 83.3%; Score 5; DB 17; Length 150;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 2 recsr 6

RESULT 13

AA82272
ID AAY82272 standard; Protein; 150 AA.

XX AAY82272;

DT 16-JUN-2000 (first entry)

DE Human bone morphogenetic protein 9 SEQ ID NO:9.

KW Bone morphogenetic protein 9; BMP-9; bone defect; cartilage defect;
KW wound healing; tissue repair; hepatic growth; antiinfectility;
KW osteopathic; antiarthritic; vulnerary.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..40
FT /Label= signal
FT 41..150
FT /Label= Bone_morphogenetic_c-protein-9

XX US6034062-A.

XX 07-MAR-2000.

PE 13-MAR-1997; 97US-0815652.

PR 13-MAR-1997; 97US-0815652.

XX (GENY) GENETICS INST INC.

XX Thies RS, Song JJ;

XX WPI; 2000-255707/22.

DR N-PSDB; AAZ95747.

XX Novel bone-morphogenetic protein (BMP)-9 polypeptides, useful in the
PT treatment of bone and cartilage defects and in wound healing and tissue
PT repair, as well as hepatic growth and function

PS Claim 1; Column 45-46; 36pp; English.

XX A method has been developed for decreasing proliferation of kidney or
CC lung epithelial cells, or for inhibiting kidney tubule formation, in a
CC patient. The method comprises administering an amount of a bone
CC morphogenetic protein 9 (BMP-9) protein. BMP-9 polypeptides from the
CC invention may be used to induce bone and cartilage formation or other
CC connective tissue formation, in periodontal disease, in wound healing
CC and tissue repair, hepatic growth and function, and in regulating lung
CC bone fractures, open fracture reduction, improved fixation of artificial
CC joints, tooth repair processes, treatment of osteoporosis and
CC osteoarthritis, and as a dietary supplement or as a component of cell
CC culture media. BMP-9 may also be used for repair and regenerating of
CC liver cells. The BMP-9 polypeptides may also have angiogenic,
CC chemotactic and/or chemottractant properties, and may induce collagen
CC synthesis, fibrosis, differentiation processes, cell proliferative
CC responses, cell adhesion responses, and migration. When dimerized, the
CC BMP-9 polypeptide is expected to demonstrate effects on the production
CC of follicle stimulating hormone (FSH), and so may be used as a
CC contraceptive, as a fertility inducing therapeutic, and for advancing
CC the onset of fertility in sexually immature mammals, so as to increase
CC the lifetime reproductive performance of domestic animals. The present
CC sequence represents human BMP-9.

SQ Sequence 150 AA;

Query Match 83.3%; Score 5; DB 21; Length 150;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 2 recsr 6

RESULT 14

AA530438
ID AAY30438 standard; Protein; 161 AA.

XX AAY30438;

DT 15-NOV-1999 (first entry)

DE Mature nematode extracted anticoagulant protein AdunAP7.

KW Nematode extracted anticoagulant protein; NAP; anticoagulant;
KW serine protease inhibitor; NAP domain; factor VIIa/TF.

XX Ancylostoma duodenale.

XX US5955294-A.

XX 21-SEP-1999.

PE 19-APR-1996; 96US-0634641.

PR 19-APR-1996; 96US-0634641.

XX 18-OCT-1994; 94US-0326110.

PR 05-JUN-1995; 95US-0461965.

PR 05-JUN-1995; 95US-0465380.

PR 05-JUN-1995; 95US-0486397.

PR 17-OCT-1995; 95US-0486399.

XX 17-OCT-1995; 95WO-US13231.

XX (CORV-) CORVUS INT INC.

XX Bergum FW, Ganssems YGJ, Jespers IS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PH;
PI Vlaek GP;

DR WPI; 1999-539569/45.

```

XX Screening an isolated protein for Nematode-extracted Anticoagulant
PT Protein domains
XX
PS Disclosure; Fig 20; 197pp; English.
XX
CC The present sequence represents a nematode extracted anticoagulant
CC protein (NAP). The protein has activity as an anticoagulant and/or serine
CC protease inhibitor. The protein contains at least one NAP domain which
CC has selective inhibitory activity for factor VIIa/TF. The specification
CC describes a method for screening an isolated protein at least one domain
CC for factor VIIa/TF selective inhibitory activity. The method comprises
CC determining the time to clotting effected by a concentration of the
CC isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo
CC activated partial thromboplastin time (APTT) assay; calculating
CC prolongation of clotting effected by the isolated protein in each of
CC the PT and APTT assay, with respect to a baseline clotting value for
CC each assay, where prolongation of clotting is calculated as fold
CC elevation of clotting time relative to a baseline clotting value, where
CC a doubling of clotting time is deemed a two-fold elevation; and
CC calculating a PR to APTT prolongation ratio, where a ratio at least
CC one is indicative of factor VIIa/TF inhibitory activity. The method is
CC useful for determining if a protein has factor VIIa/TF inhibitory
CC activity.
XX
SQ Sequence 161 AA;

Query Match 83.3%; Score 5; DB 20; Length 161;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RECSR 6
   |||||
DB 43 recsr 47

RESULT 15
AAV30436
ID AAV30436 standard; Protein: 162 AA.
AC
XX AAV30436;
XX
DT 15-NOV-1999 (first entry)
XX
DE Mature nematode extracted anticoagulant protein AcanNAP45.
XX
KM Nematode extracted anticoagulant protein; NAP; anticoagulant;
XX serine protease inhibitor; NAP domain; factor VIIa/TF.
OS Ancylostoma caninum.
XX
XX US5955294-A.
XX
XX 21-SEP-1999.
XX
XX 19-APR-1996; 96US-0634641.
XX
XX 19-APR-1996; 96US-0634641.
XX
XX 18-OCT-1994; 94US-0326110.
XX
XX 05-JUN-1995; 95US-0461965.
XX
XX 05-JUN-1995; 95US-0465380.
XX
XX 05-JUN-1995; 95US-0486397.
XX
XX 05-JUN-1995; 95US-0486399.
XX
XX 17-OCT-1995; 95WO-US13231.
XX
XX (CORV-) CORVAS INT INC.
XX
PI Bergum PW, Gansemans YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PEH;
PI Vlasuk GP;
XX
DR WPI; 1999-539569/45.

```

```

XX Screening an isolated protein for Nematode-extracted Anticoagulant
PT Protein domains
XX
PS Disclosure; Fig 18; 197pp; English.
XX
CC The present sequence represents a nematode extracted anticoagulant
CC protein (NAP). The protein has activity as an anticoagulant and/or serine
CC protease inhibitor. The protein contains at least one NAP domain which
CC has selective inhibitory activity for factor VIIa/TF. The specification
CC describes a method for screening an isolated protein at least one domain
CC for factor VIIa/TF selective inhibitory activity. The method comprises
CC determining the time to clotting effected by a concentration of the
CC isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo
CC activated partial thromboplastin time (APTT) assay; calculating
CC prolongation of clotting effected by the isolated protein in each of
CC the PT and APTT assay, with respect to a baseline clotting value for
CC each assay, where prolongation of clotting is calculated as fold
CC elevation of clotting time relative to a baseline clotting value, where
CC a doubling of clotting time is deemed a two-fold elevation; and
CC calculating a PR to APTT prolongation ratio, where a ratio at least
CC one is indicative of factor VIIa/TF inhibitory activity. The method is
CC useful for determining if a protein has factor VIIa/TF inhibitory
CC activity.
XX
SQ Sequence 162 AA;

Query Match 83.3%; Score 5; DB 20; Length 162;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RECSR 6
   |||||
DB 44 recsr 48

```

Search completed: February 26, 2002, 01:50:49
Job time: 1253 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2002, 23:59:25 ; Search time 62.13 seconds
(without alignments)
196.484 Million cell updates/sec

Title: US-09-819-136-2
Perfect score: 3016
Sequence: 1 MPALRPLPLPLLLRLTSGA.....KRLLELLEKQACELNRFQD 548

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	7.1	143	2	US-08-422-333-10
2	213	7.1	143	6	5223482-20
3	213	7.1	144	6	5187153-18
4	213	7.1	147	1	US-08-358-160-72
5	210	7.0	123	6	5466783-21
6	208	6.9	123	6	5466783-21
7	207	6.9	122	2	US-08-422-333-12
8	207	6.9	122	6	5187153-20
9	207	6.9	122	6	5220013-23
10	207	6.9	127	6	5466783-24
11	206	6.8	122	6	5223482-22
12	200.5	6.6	122	6	5466783-23
13	196	6.5	213	6	5466783-25
14	194.5	6.4	276	1	US-07-828-920A-1
15	194.5	6.4	276	1	US-08-437-841-9
16	194.5	6.4	276	1	US-08-286-521-9
17	194.5	6.4	276	1	US-08-438-175-9
18	194.5	6.4	276	2	US-08-796-850-1
19	194.5	6.4	276	4	US-08-854-764-3
20	194.5	6.4	276	4	US-08-943-682-9
21	194.5	6.4	276	5	PCT-US95-09377-3
22	194.5	6.4	276	5	PCT-US95-09464-9
23	194.5	6.4	277	1	US-07-844-297-1
24	194.5	6.4	304	1	US-08-026-145-2
25	194.5	6.4	304	1	US-08-446-646-9
26	194.5	6.4	304	1	US-08-676-125A-18
27	194.5	6.4	304	2	US-09-136-012A-18

28	194.5	6.4	304	3	US-08-676-124-1	Sequence 1, Appl1
29	194.5	6.4	304	3	US-08-208-264A-25	Sequence 25, Appl1
30	194.5	6.4	304	3	US-09-414-878-1	Sequence 1, Appl1
31	194.5	6.4	304	4	US-09-240-136-1	Sequence 1, Appl1
32	194.5	6.4	304	4	US-09-054-782-2	Sequence 2, Appl1
33	194.5	6.4	304	6	5466783-2	Patent No. 5466783
34	194.5	6.4	352	3	US-08-854-764-2	Sequence 2, Appl1
35	194.5	6.4	352	5	PCT-US95-09377-2	Sequence 2, Appl1
36	192	6.4	252	1	US-08-685-660A-7	Sequence 7, Appl1
37	192	6.4	252	2	US-08-974-196-7	Sequence 7, Appl1
38	192	6.4	252	4	US-09-071-709-10	Sequence 10, Appl1
39	192	6.4	252	4	US-09-013-896A-2	Sequence 2, Appl1
40	179.5	6.0	235	1	US-08-147-710-2	Sequence 2, Appl1
41	179.5	6.0	235	1	US-08-458-090-2	Sequence 2, Appl1
42	179.5	6.0	235	2	US-08-457-887-2	Sequence 2, Appl1
43	179.5	6.0	235	3	US-08-817-145-3	Sequence 3, Appl1
44	179	5.9	213	2	US-08-796-850-2	Sequence 2, Appl1
45	179	5.9	1940	2	US-08-644-271-30	Sequence 30, Appl1

ALIGNMENTS

```
RESULT 1
US-08-422-333-10
; Sequence 10, Application US/0842233
; Patent No. 5912410
; GENERAL INFORMATION:
; APPLICANT: CORDELL, Barbara L.
; TITLE OF INVENTION: TRANSGENIC NON-HUMAN MAMMAL DISPLAYING
; TITLE OF INVENTION: THE AYLLOID-FORMING PATHOLOGY OF ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scios, Inc.
; STREET: 2450 Bayshore Parkway
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,333
; FILING DATE: 13-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Shearer, Peter R.
; REGISTRATION NUMBER: 28,117
; REFERENCE/DOCKET NUMBER: 21900-28048.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 966-1550
; TELEFAX: (415) 966-2438
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 143 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-422-333-10

Query Match 7.1%; Score 213; DB 2; Length 143;
Best Local Similarity 39.3%; Pred. No. 2.3e-07;
Matches 42; Conservative 13; Mismatches 46; Indels 6; Gaps 2;

QY 306 CIGPISPHLVLMHYDQRCGCTFPARCGDGAARGETYEACQACARPGDACYLPVAVQ 365
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 35 CAGMNSRYF-----YNGTSNACETFOYGGCGMGNGNNEVTEKEIDQTCRTVA--ACNLPIVR 88
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 366 GPCRGMEPRMAYSPLLQCHPFVYGGCEGNGNNEFHRESCEDACPVF 412
```

Db 89 GPCRAFIQLMAFDAVKGKCVLEFPYGGCGGNGNKFYSEKREYCGVP 135

RESULT 2
5223482-20
; Patent No. 5223482
; APPLICANT: SCHILLING, JAMES W.; PONTE, PHYLLIS A.; CORDELL,
; BARBARA
; TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
; INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/361,912
; FILING DATE: 06-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 359,911
; FILING DATE: 12-MAY-1989
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO: 20:
; LENGTH: 143
5223482-20

Query Match 7.1%; Score 213; DB 6; Length 143;
Best Local Similarity 39.3%; Pred. No. 2.3e-07;
Matches 42; Conservative 13; Mismatches 46; Indels 6; Gaps 2;

OY 306 CTGPTSPHLVLMHYDPORGCMTPPARCGDGAARGFETYEACQACARGDGVLPAYQ 365
Db 35 CMGMSRYF---YNGTSMACETFOYGGCMGNGNMFYTERKECLQCTCTVA--ACNLPVIR 88
OY 366 GPCGWERMAVSPLLQCCHFYVGGCGEGNGNNHSHSCECEDACPVP 412
Db 89 GPCRAFIQLMAFDAVKGKCVLEFPYGGCGGNGNKFYSEKREYCGVP 135

RESULT 3
5187153-18
; Patent No. 5187153
; APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
; AMYLOID POLYPEPTIDE DERIVATIVES
; NUMBER OF SEQUENCES: 33
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/502,273
; FILING DATE: 29-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 361,912
; FILING DATE: 06-JUN-1989
; APPLICATION NUMBER: 359,911
; FILING DATE: 12-MAY-1989
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO: 18:
; LENGTH: 144
5187153-18

Query Match 7.1%; Score 213; DB 6; Length 144;

Best Local Similarity 39.3%; Pred. No. 2.3e-07;
Matches 42; Conservative 13; Mismatches 46; Indels 6; Gaps 2;

OY 306 CTGPTSPHLVLMHYDPORGCMTPPARCGDGAARGFETYEACQACARGDGVLPAYQ 365
Db 35 CMGMSRYF---YNGTSMACETFOYGGCMGNGNMFYTERKECLQCTCTVA--ACNLPVIR 88
OY 366 GPCGWERMAVSPLLQCCHFYVGGCGEGNGNNHSHSCECEDACPVP 412
Db 89 GPCRAFIQLMAFDAVKGKCVLEFPYGGCGGNGNKFYSEKREYCGVP 135

RESULT 4
US-08-358-160-72
; Sequence 72; Application US/08358160
; Patent No. 5663143
; GENERAL INFORMATION:
; APPLICANT: LEY, Arthur C.
; APPLICANT: LADNER, Robert C.
; APPLICANT: GUTERMAN, Sonia K.
; APPLICANT: ROBERTS, Bruce L.
; APPLICANT: MARKLAND, William
; APPLICANT: KENT, Rachel B.
; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
; NUMBER OF SEQUENCES: 234
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W. Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,160
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/133,031
; FILING DATE: 13-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/009,319
; FILING DATE: 26-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/664,989
; FILING DATE: 01-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/487,063
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/240,160
; FILING DATE: 02-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 147 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-358-160-72

Query Match 7.1%; Score 213; DB 1; Length 147;

Best Local Similarity 40.2%; Pred. No. 2.4e-07; Indels 6; Gaps 2;

Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

QY 306 GPTSPHLVLMHYDPORGCMTFPAAGCDGAARGFETYEACQACARPGDACPVP 365

DB 35 CMGMSRYF---YNGTSMACETFOYGGCMGNGNMFVTEKECLQCTCTVA--ACNLPYVR 88

QY 366 GPCRGWRWVAYSPLLQOCHPFYVGGCEGNGNMFHRSRESCEDACPVP 412

DB 89 GPCRATQLMAFDVAQKCVLFYGGCGNGNMFYSQKECKEYCGVP 135

RESULT 5

5466783-22

; Patent No. 5466783

; APPLICANT: Wun, Tze-Cheln, Kretzmer, Kuniko K.; Broze,

; George J. Jr.

; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR

; NUMBER OF SEQUENCES: 26

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/93,285

; FILING DATE: 15-JUL-1993

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 566,280

; FILING DATE: 13-AUG-1990

; APPLICATION NUMBER: 123,753

; FILING DATE: 23-NOV-1987

; APPLICATION NUMBER: 77,366

; FILING DATE: 23-JUL-1987

; SEQ ID NO:22:

; LENGTH: 123

5466783-22

Query Match 7.0%; Score 210; DB 6; Length 123;

Best Local Similarity 39.0%; Pred. No. 3.1e-07; Indels 2; Gaps 1;

Matches 41; Conservative 14; Mismatches 48; Indels 2; Gaps 1;

QY 308 GPTSPHLVLMHYDPORGCMTFPAAGCDGAARGFETYEACQACARPGDACPVP 367

DB 12 GPCLGMIKRFYNGSSMACETFFHYGCMGNGNMFVSKELQTCR--TVSACSLPIYQGP 69

QY 368 CRGWPRWVAYSPLLQOCHPFYVGGCEGNGNMFHRSRESCEDACPVP 412

DB 70 CRAFTILMAFDVAQKCVLFYGGCGNGNMFYSQKECKEYCGVP 114

RESULT 6

5466783-21

; Patent No. 5466783

; APPLICANT: Wun, Tze-Cheln, Kretzmer, Kuniko K.; Broze,

; George J. Jr.

; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR

; NUMBER OF SEQUENCES: 26

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/93,285

; FILING DATE: 15-JUL-1993

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 566,280

; FILING DATE: 13-AUG-1990

; APPLICATION NUMBER: 123,753

; FILING DATE: 23-NOV-1987

; APPLICATION NUMBER: 77,366

; FILING DATE: 23-JUL-1987

; SEQ ID NO:21:

; LENGTH: 123

5466783-21

Best Local Similarity 38.1%; Pred. No. 4.2e-07; Indels 2; Gaps 1;

Matches 40; Conservative 14; Mismatches 49; Indels 2; Gaps 1;

QY 308 GPTSPHLVLMHYDPORGCMTFPAAGCDGAARGFETYEACQACARPGDACPVP 367

DB 12 GPCLGMIKRFYNGTSMACETFFHYGCMGNGNMFVTEKECLQCTCTVA--ACNLPYVR 69

QY 368 CRGWPRWVAYSPLLQOCHPFYVGGCEGNGNMFHRSRESCEDACPVP 412

DB 70 CRAFTILMAFDVAQKCVLFYGGCGNGNMFYSQKECKEYCGVP 114

RESULT 7

US-08-422-333-12

; Sequence 12, Application US/08422333

; Patent No. 5912410

; GENERAL INFORMATION:

; APPLICANT: CORDELL, Barbara L.

; TITLE OF INVENTION: TRANSGENIC NON-HUMAN MAMMAL DISPLAYING

; THE AMYLOID-FORMING PATHOLOGY OF ALZHEIMER'S DISEASE

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Scios, Inc.

; STREET: 2450 Bayshore Parkway

; CITY: Mountain View

; STATE: CA

; COUNTRY: USA

; ZIP: 94043

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/422,333

; FILING DATE: 13-APR-1995

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Shearer, Peter R.

; REGISTRATION NUMBER: 28,117

; REFERENCE/DOCKET NUMBER: 21900-28048.00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 966-1550

; TELEFAX: (415) 968-2438

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 122 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-422-333-12

Query Match 6.9%; Score 207; DB 2; Length 122;

Best Local Similarity 38.1%; Pred. No. 4.9e-07; Indels 2; Gaps 1;

Matches 40; Conservative 16; Mismatches 47; Indels 2; Gaps 1;

QY 308 GPTSPHLVLMHYDPORGCMTFPAAGCDGAARGFETYEACQACARPGDACPVP 367

DB 12 GPCLGIFKRFYNGTSMACETFFHYGCMGNGNMFVTEKECLQCTCTVA--TVSACSLPIYQGP 69

QY 368 CRGWPRWVAYSPLLQOCHPFYVGGCEGNGNMFHRSRESCEDACPVP 412

DB 70 CRAFTILMAFDVAQKCVLFYGGCGNGNMFYSQKECKEYCGVP 114

RESULT 8

5187153-20

; Patent No. 5187153

; APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO

; TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S

; AMYLOID POLYPEPTIDE DERIVATIVES

; NUMBER OF SEQUENCES: 33

Query Match 6.9%; Score 208; DB 6; Length 123;

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/502,273
FILING DATE: 29-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 361,912
FILING DATE: 06-JUN-1989
APPLICATION NUMBER: 359,911
FILING DATE: 12-MAY-1989
APPLICATION NUMBER: 87,002
FILING DATE: 18-AUG-1987
APPLICATION NUMBER: 8,810
FILING DATE: 30-JAN-1987
APPLICATION NUMBER: 948,376
FILING DATE: 31-DEC-1986
APPLICATION NUMBER: 932,193
FILING DATE: 17-NOV-1986
SEQ ID NO:20:
LENGTH: 122
5187153-20

Query Match 6.9%; Score 207; DB 6; Length 122;
Best Local Similarity 38.1%; Pred. No. 4.9e-07;
Matches 40; Conservative 16; Mismatches 47; Indels 2; Gaps 1;
QY 308 GPTSPHLVIMHYDPORGCGMTFPARCGDGAARGFETYEACQACARPGDACLPAVQGP 367
DB 12 GPCGLGKFRFYNGTSMACETFLYGGCMGNLNNFLSOKECLOTCTCR--TVEACNLPYVQGP 69
QY 368 CRGMEPRMAYSPILQOCHPFYVYGCGEGNGNNFHSRESCDACPVP 412
DB 70 GCAFIQIUMAFDAVKGKCVRFSGCGKNGNKFYSQKCKEYCGIP 114

RESULT 9
5220013-23
PATENT NO. 5220013
APPLICANT: PONTE, PHYLLIS A.;CORDELL, BARBARA
TITLE OF INVENTION: DNA SEQUENCE USEFUL FOR THE DETECTION
OF ALZHEIMER S DISEASE
NUMBER OF SEQUENCES: 30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/444,118
FILING DATE: 30-NOV-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 87,002
FILING DATE: 18-AUG-1987
APPLICATION NUMBER: 8,810
FILING DATE: 30-JAN-1987
APPLICATION NUMBER: 948,376
FILING DATE: 31-DEC-1986
APPLICATION NUMBER: 932,193
FILING DATE: 17-NOV-1986
SEQ ID NO:23:
LENGTH: 122
5220013-23

Query Match 6.9%; Score 207; DB 6; Length 122;
Best Local Similarity 38.1%; Pred. No. 4.9e-07;
Matches 40; Conservative 16; Mismatches 47; Indels 2; Gaps 1;
QY 308 GPTSPHLVIMHYDPORGCGMTFPARCGDGAARGFETYEACQACARPGDACLPAVQGP 367
DB 12 GPCGLGKFRFYNGTSMACETFLYGGCMGNLNNFLSOKECLOTCTCR--TVEACNLPYVQGP 69
QY 368 CRGMEPRMAYSPILQOCHPFYVYGCGEGNGNNFHSRESCDACPVP 412
DB 70 GCAFIQIUMAFDAVKGKCVRFSGCGKNGNKFYSQKCKEYCGIP 114

RESULT 10
5466783-24

PATENT NO. 5466783
APPLICANT: WUN, TZE-CHIEIN.;KREITZMER, KUNIKO K.;BROZE,
GEORGE J. JR.
TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
NUMBER OF SEQUENCES: 26
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/93,285
FILING DATE: 15-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 566,280
FILING DATE: 13-AUG-1990
APPLICATION NUMBER: 123,753
FILING DATE: 23-NOV-1987
APPLICATION NUMBER: 77,366
FILING DATE: 23-JUL-1987
SEQ ID NO:24:
LENGTH: 127
5466783-24

Query Match 6.9%; Score 207; DB 6; Length 127;
Best Local Similarity 40.4%; Pred. No. 5.1e-07;
Matches 44; Conservative 12; Mismatches 45; Indels 8; Gaps 3;
QY 306 CTGPTSPHLVIMHYDPORGCGMTFPARCGDGAARGFETYEACQACARPGDACLPAVQ 365
DB 14 CMGMSIRFE----NGTSMACETFOYGGCMGNNGNNFYTEKECLOTCTRTVA--ACNLPYVR 67
QY 366 CRGMEPRMAYSPILQOCHPFYVYGCGEGNGNNFHSRESC--DACPVP 412
DB 68 GPCRAFIQIUMAFDAVKGKCVLPYGGCGGNGNKFYSQKCKEYCGIP 116

RESULT 11
5223482-22
PATENT NO. 5223482
APPLICANT: SCHILLING, JAMES W.;PONTE, PHYLLIS A.;CORDELL,
BARBARA
TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
NUMBER OF SEQUENCES: 34
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/361,912
FILING DATE: 06-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 359,911
FILING DATE: 12-MAY-1989
APPLICATION NUMBER: 87,002
FILING DATE: 18-AUG-1987
APPLICATION NUMBER: 8,810
FILING DATE: 30-JAN-1987
APPLICATION NUMBER: 948,376
FILING DATE: 31-DEC-1986
APPLICATION NUMBER: 932,193
FILING DATE: 17-NOV-1986
SEQ ID NO:22:
LENGTH: 122
5223482-22

Query Match 6.8%; Score 206; DB 6; Length 122;
Best Local Similarity 39.6%; Pred. No. 5.7e-07;
Matches 38; Conservative 16; Mismatches 40; Indels 2; Gaps 1;
QY 317 WHYDPORGCGMTFPARCGDGAARGFETYEACQACARPGDACLPAVQGPCRGMEPRMA 376
DB 23 YFNGTSMACETFLYGGCMGNLNNFLSOKECLOTCTCR--TVEACNLPYVQGPCRAFIQIUMA 80
QY 377 YSPILQOCHPFYVYGCGEGNGNNFHSRESCDACPVP 412
DB 81 FDAVKGKCVRFSGCGKNGNKFYSQKCKEYCGIP 116

RESULT 12
5466783-23
; Patent No. 5466783
; APPLICANT: Wun, Tze-Cheln, Kretzmer, Kunlko K., Broze,
; George J. Jr.
; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/93,285
; FILING DATE: 15-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 566,280
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 123,753
; FILING DATE: 23-NOV-1987
; APPLICATION NUMBER: 77,366
; FILING DATE: 23-JUL-1987
; SEQ ID NO: 23:
; LENGTH: 122
5466783-23

Query Match 6.6%; Score 200.5; DB 6; Length 122;
Best Local Similarity 38.1%; Pred. No. 1.3e-06;
Matches 40; Conservative 16; Mismatches 46; Indels 3; Gaps 2;
QY 308 GPTSPHLVLMHYDPQGGCMTEPPARGCDGAARGFETYEACQACARPGDACLPAVQGP 367
DB 12 GCLGLFKRYFYNGTSMACETE-LGCMGNLNNFLSQKECIQTGR--TVEACNLPIVQGP 68
QY 368 CRGMEPRMAYSPILQOCHPFYVYGGCEGNGNNFHRSCECDACPRV 412
DB 69 CRAFTQMAFDVAVKGVAFSTYSGCKGNKNTFSQKECYCGIP 113

RESULT 13
5466783-25
; Patent No. 5466783
; APPLICANT: Wun, Tze-Cheln, Kretzmer, Kunlko K., Broze,
; George J. Jr.
; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/93,285
; FILING DATE: 15-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 566,280
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 123,753
; FILING DATE: 23-NOV-1987
; APPLICATION NUMBER: 77,366
; FILING DATE: 23-JUL-1987
; SEQ ID NO: 25:
; LENGTH: 213
5466783-25

Query Match 6.5%; Score 196; DB 6; Length 213;
Best Local Similarity 30.2%; Pred. No. 4.6e-06;
Matches 49; Conservative 20; Mismatches 59; Indels 34; Gaps 5;
QY 263 CTARNAAGILRADPFLSVQREPARDAAPSIPAPACL--PVOACTGPTSPHLVLMHYD 320
DB 58 CTRDANARIK-----TTLQOE-----KPDGCFLEDPGICRG---YITRYFYN 98
QY 321 PORGCMTFPARGCDGAARGFETYEACQACARPG-----DACVLPAVQGP 367
DB 99 NOTKCCERFKYGGCLCNMNNFETLECKNICEDGPNQFOVDNYGFERHGPSWCLTPADKGL 158
QY 368 CRGMEPRMAYSPILQOCHPFYVYGGCEGNGNNFHRSCECDAC 409
DB 159 CRANENRFYNSVIGKCRPFKYSGCCGNMNNFTSKQECILRAC 200

RESULT 14
US-07-828-920A-1
; Sequence 1, Application US/07828920A
; Patent No. 5312736
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Jesper
; APPLICANT: No. 5312736dfang, Ole Juul
; TITLE OF INVENTION: Anticoagulant Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 53127360 No. 5312736disk of No. 5312736th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6200
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/828,920A
; FILING DATE: 19920127
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 4080/89
; FILING DATE: 18-AUG-1989
; APPLICATION NUMBER: WO PCT/DK90/00212
; FILING DATE: 17/AUG/1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T.
; REGISTRATION NUMBER: 30335
; REFERENCE/DOCKET NUMBER: 3287,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 276 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..276
US-07-828-920A-1

Query Match 6.4%; Score 194.5; DB 1; Length 276;
Best Local Similarity 27.3%; Pred. No. 7.6e-06;
Matches 50; Conservative 20; Mismatches 58; Indels 55; Gaps 5;
QY 263 CTARNAAGILRADPFLSVQREPARDAAPSIPAPACL--PVOACTGPTSPHLVLMHYD 320
DB 76 CTRDANARIK-----TTLQOE-----KPDGCFLEDPGICRG---YITRYFYN 116
QY 321 PORGCMTFPARGCDGAARGFETYEACQACARPG-----DACVLPAVQGP 367
DB 117 NOTKCCERFKYGGCLCNMNNFETLECKNICEDGPNQFOVDNYGTLNANNSTLPQSTK 176
QY 357 -----DACVLPAVQGPGRMEPRMAYSPILQOCHPFYVYGGCEGNGNNFHRSCE 406
DB 177 VPSLEFHPGPSWCLTPADKGLCRANENRFYNSVIGKCRPFKYSGCCGNMNNFTSKQECIL 236
QY 407 DAC 409
DB 237 RAC 239

RESULT 15
US-08-437-841-9
; Sequence 9, Application us/08437841
; Patent No. 5563123
; GENERAL INFORMATION:
; APPLICANT: Innis, Michael
; TITLE OF INVENTION: Chimeric Proteins
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton St.
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,841
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,521
; FILING DATE: 05-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saveriede, Paul B.
; REGISTRATION NUMBER: 36,914
; REFERENCE/DOCKET NUMBER: 0990.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2585
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 276 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-437-841-9

Query Match 6.4%; Score 194.5; DB 1; Length 276;
Best Local Similarity 27.3%; Pred. No. 7.6e-06;
Matches 50; Conservative 20; Mismatches 58; Indels 55; Gaps 5;
QY 263 CTARNAAGLRADPLSVVOREPARDAAPSIAPAECL--PDVQACTGPTSPHLVLMHYD 320
Db 76 CTRDNANRIK-----TTLQOE-----KPDFCFLEEDPGICNG---YITRYFYV 116
QY 321 PORGCMTFPARGCDGAARGFETYEACQACARPG----- 356
Db 117 NQTKOCERFHYGGGLGMNNFETLECKNICEDGPNGFQVDNYGTQLNAVNSLTLPQSTR 176
QY 357 -----DACVLPVQGPCRGMEPRMAYSPLLQOCHPFYVGGCGNGNNHRSRSC 406
Db 177 VPSLFEFHGSPWCLTPADRLCRANENRFYNSVIGKCRPFKYSGCGMNNFTSKQCL 236
QY 407 DAC 409
Db 237 RAC 239

Search completed: February 26, 2002, 01:10:16
Job time: 4251 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:10:10 ; Search time 118.08 seconds
(without alignments)
678.838 Million cell updates/sec

Title: US-09-819-136-2
Perfect score: 3016
Sequence: 1 MPALRPULPLLLRLTSGA.....KKILELLEKQACELNRFQD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPTRMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294.5	9.8	3198	5 Q9U8G8	Q9U8G8 manduca sex
2	260	8.6	3060	5 Q9VAV4	Q9VAV4 drosophila
3	259	8.6	1572	5 Q44938	Q44938 haemochus
4	251.5	8.3	2167	5 Q76840	Q76840 caenorhabd1
5	251	8.3	2174	5 Q9G8R0	Q9G8R0 drosophila
6	236.5	7.8	2225	5 Q45881	Q45881 caenorhabd1
7	218	7.2	1043	5 Q17644	Q17644 caenorhabd1
8	216.5	7.2	144	11 Q9QW87	Q9QW87 mesocricetu
9	216	7.2	922	5 Q21418	Q21418 caenorhabd1
10	213	7.1	151	4 P78491	P78491 homo sapien
11	203	6.7	349	11 Q9DBJ9	Q9DBJ9 mus musculu
12	202.5	6.7	1743	5 Q9XW5	Q9XW5 caenorhabd1
13	196	6.5	352	11 Q70160	Q70160 cavia porce
14	195	6.5	1965	5 Q61893	Q61893 caenorhabd1
15	194.5	6.4	287	13 Q93424	Q93424 cyprianus ca
16	194.5	6.4	2150	5 Q44131	Q44131 caenorhabd1
17	193.5	6.4	1235	5 Q95428	Q95428 homo sapien
18	193	6.4	4123	4 Q75851	Q75851 homo sapien
19	191.5	6.3	1474	5 Q62504	Q62504 caenorhabd1

20	188.5	6.2	1280	11 Q9EPX2	Q9EPX2 mus musculu
21	184.5	6.1	246	11 Q9Z2U8	Q9Z2U8 mus musculu
22	183.5	6.1	342	13 P70004	P70004 xenopus lae
23	182	6.0	197	5 P82968	P82968 mellitbae c
24	181.5	6.0	2026	4 Q00468	Q00468 homo sapien
25	179.5	6.0	1208	5 Q23456	Q23456 caenorhabd1
26	179	5.9	1599	5 Q09983	Q09983 caenorhabd1
27	177	5.9	763	5 Q9XZD0	Q9XZD0 drosophila
28	175	5.8	396	6 Q28874	Q28874 canis famli
29	173.5	5.8	251	4 Q95103	Q95103 homo sapien
30	173	5.7	1549	11 Q60444	Q60444 cricetus
31	172	5.7	685	6 Q9TTS5	Q9TTS5 bos taurus
32	172	5.7	1344	11 Q9Z214	Q9Z214 mus musculu
33	171.5	5.7	947	5 Q44171	Q44171 caenorhabd1
34	171.5	5.7	947	5 Q26262	Q26262 caenorhabd1
35	170.5	5.7	838	5 Q27422	Q27422 caenorhabd1
36	170.5	5.7	838	5 Q18761	Q18761 caenorhabd1
37	170.5	5.7	1419	13 Q98SW3	Q98SW3 brachydanio
38	170	5.6	67	11 Q9QW86	Q9QW86 mesocricetu
39	169	5.6	1391	5 Q19021	Q19021 caenorhabd1
40	168.5	5.6	1395	5 Q44924	Q44924 drosophila
41	167.5	5.6	273	4 Q9WTP5	Q9WTP5 homo sapien
42	166	5.5	1511	5 Q9VB21	Q9VB21 drosophila
43	165.5	5.5	1395	5 Q9W213	Q9W213 drosophila
44	164.5	5.5	1273	5 Q44928	Q44928 caenorhabd1
45	163	5.4	1160	5 Q18261	Q18261 caenorhabd1

ALIGNMENTS

RESULT 1
Q9U8G8 PRELIMINARY: PRT: 3198 AA.
AC Q9U8G8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LACUNIN PRECURSOR.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99457716; PubMed=10528409;
RA Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;
RT "Expression of lacunin, a large multidomain extracellular matrix
protein, accompanies morphogenesis of epithelial monolayers in Manduca
sexta.";
RT Insect Biochem. Mol. Biol. 29:883-897(1999).
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC -!- SIMILARITY: TO DOMAIN.
CC DOMAIN.
CC -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
CC EMBL: AF078161; AAFA04457.1; .
DR HSSP: P12111; ZKNT.
DR InterPro: IPR003598; IG_C2.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR002221; WAP.
DR Pfam: PF00047; IG_2.
DR Pfam: PF00014; Kunitz_BPTI; 9.
DR Pfam: PF00095; WAP; 1.
DR PRINTS: PR00759; BASICPRASE.
DR SMART: SM00408; IGC2; 2.
DR SMART: SM00131; KU; 10.
DR SMART: SM00209; TSP1; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 8.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 10.

Query Match 9.8%; Score 294.5; DB 5; Length 3198;
 Best Local Similarity 26.2%; Pred. No. 9.4e-15;
 Matches 113; Conservative 40; Mismatches 171; Indels 107; Gaps 19

```

01 35 NOLSPNLNMDAOSTCTCRECSRSDCCAAAEKCCINVCGLHSCVAARFPFSPAAPTAAACE 94
02 1709 NIISSSTIAENCTSTSEFGCCFQJDETPA-----TGPQGGCCCNSTBERGCCP 175
03 95 GFVCPQOOGSDCIDMDOPVRCRDRCEKSEPSFTCASGILTYNR-----CYNDAEACHR 148
04 1755 DGVSFARGND-----FEG-----CYTICNLIS-SYGGCCPGETPARGPDJLGCILLSAVGCCP 180
05 149 -----GLAHIVPCKHVLSWMPSSBP--PPTTARPTGGAAPVPALYSSSPQANOV 199
06 1806 DNRKPAVGFHLEGCCQCY-----SFGCCCPDNATVARGN-----FQGGCGQCYEH 185
07 200 GGTASLHCDVSGRP-----PRAVYWEK-QSHQRENLIMRPMOMTGNVYVS 244
08 1852 GCCCPRHTEAAGPPYDGGCGCTHYGCGCCPGVATIAKGQNHQGCDC--RDSQYG-----C 1903
09 245 IGQLVLYNAREPDEGLYTCYARNA-----GGLRA-----DFPLSYQREPPARDAAPS 292
10 1904 CGDGGTPTATGFEREGECDOANSEVGGCCPDGLTEKHKHFTGCSLDLEN-----KQAAAG 1956
11 293 IPAPAECLPDVQACTGTSPLHLYLMHYDLPORGCMTPFARGCDGAARFETYEACQACA 352
12 1957 LPHD-----RGTCRMYSVYMYWYDLEYEGGCSFRFYGGCEGNGNFFAIKRECEDEVCV 2006
13 353 R-GFGDCAVLEPAVOGPRGMEPRMAYSPLLLOCHPRVYVGGCGGNGNNHSHSECEDACPV 411
14 2007 QPAEKDCAINLPKVKAGACIGTINIMFYADAQEQSCQFTYGGCIGMANNVASTLQEOECOP 2066
15 412 PRTPPCNACRL 422
16 2067 ERSED--QCRL 2075

```

RESULT	2	
09VAV4		
ID	09VAV4	PRELIMINARY;
AC	09VAV4; 09VAV3;	PRT; 3060 AA.
DT	01-MAY-2000 (TREMblrel. 13, Created)	
DT	01-JUN-2001 (TREMblrel. 16, Last sequence update)	
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)	
DE	CG1540 PROTEIN.	
GN	PPN OR CG1540 OR CG18436.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha	
OX	Ephydroidea; Drosophilidae; Drosophila.	
RN	NCBI_TaxId=7227;	
RP	[1]	
RC	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132.	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	Georges R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers J.R., Vardell S., Blazer R.G., Champe M., Pfeiffer B.D.,	
RA	Wan R.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Apoll J.F., Agapayni A., An H.-J., Andrews-Neelson C.R., Baldwin D.,	
RA	Bailly R.M., Basu A., Baxendale U., Bayraktaroglu L., Beasley E.M.,	
RA	Borkova K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Botchan M.R., Bouck J., Brokstein P., Brothier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	

RA da Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dosson K., Doup L.E., Downes M., Diugan-Rocha S., Dunkov B.C., Dunn P.
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C.A., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapten C.H., Ke Z., Kennison J.A., Ketchum R.A.,
RA Kamel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Mosherfi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Spier E.C., Sidon-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Switzer R., Teclor C., Turner C., Turner R., Venter E., Sun E.,
Wang Z.-Y., Weissman D.A., Woodley K.C., Wu D., Yang S., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
ye J., Yeh R.-F., Yaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
CC EMBL: AE003765; AA56794.2; -;
DR EMBL: AE003765; AA56795.2; -;
DR HSSP: P12111.2KNT1;
DR FlyBase: FBgn000313; Pnu.
DR InterPro: IPRO00561; EGF-like.
DR InterPro: IPRO03598; Ig_c2.
DR InterPro: IPRO03006; IG_MHC.
DR InterPro: IPRO02223; Kunitz_BPTI.
DR InterPro: IPRO00884; TSPL.
DR InterPro: IPRO02221; WAP.
DR Pfam: PF00047; Iy_3.
DR Pfam: PF00014; Kunitz-BPTI; 12.
DR Pfam: PF00090; tsp_1; 5.
DR Pfam: PF00095; wap; 1.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2; 3.
DR SMART: SM00131; KU; 12.
DR SMART: SM00209; TSPL; 7.
DR SMART: SM0217; WAP; 1.
DR PROSITE: PS00317; wap; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 11.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 12.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS50092; TSPL; 3.
KW Alternative splicing; Serine protease inhibitor.
FT VASPPlic 2803 2803
FT FT 2843 2854
FT VASPPlic 2843 2854
FT VASPPlic 2855 3060
SEQUENCE 3060 AA; 331579 MW; ACQADJ3EE58C7CO CRC64;

Query Match	8.6%	Score 260;	DB 5;	Length 3060;
Best Local Similarity	21.0%	Pred. No. 5,3e-12;		
Matches 119;	Conservative 41;	Mismatches 212;	Indels 196;	Gaps 24;
QY 18	SGAG----	LLPGIGSHPGVCPNQLSPNL-----	WYDAOSTGER-----	52
Db 1781	SGCGGNKNPFLESCEDHCPRQVAKDICEIPAEVGECACTVMSWITDQDCKRQFYFG			1840
QY 53	-GSRDDCCAAAEKCCINNVGILSHCSVAALFPSPSPALPTTASGEGFVCPDQGGSCDIDW---			108

```

Db 1841 GCGGNNRPTEESCLARCDKPEPTTTTPATRPQPSRODVCEDEBP--GECSTWLK 1897
QY 109 --DGQVPCRCRCKEKEPFTCASDGLTYNNRC-----YMAEACLRHLHIYVCKHV 160
Db 1898 WHFDRIIGACRO-----FYGNCGGNGNRFETENDCQ-----QRC 1932
QY 161 LSMPPSSPPPTTARPRGAAPVPP-ALYSPPS-----QAVGV 199
Db 1933 LSGEPAPAPPRAPAP-PTQPPDPAPIVACSQPADPGCDKALHMNYETEGRCQSFY 1991
QY 200 GG-----TASLHCDVS-----GRPPAVTWKQSHOREN--LIMRPDMYGNV 240
Db 1992 GCGCGNDNRFATEECSACSCSINIDIRIGADP-----VEHDTSCFLAFEPGNCYNNV 2044
QY 241 -----VYTSIGQLVLYNARPEDA-----GLYTCARNAAGLL----- 272
Db 2045 TMEFTYNSAEGLCDEFYVYTCGGNANNVATEECCONECNDQOTTCALPYVRGCSDSLRRW 2104
QY 273 -----RADFP-----LSVQREPARDAAPSIAPAECL--PDVQA 305
Db 2105 YPDERSGEHEFEFTGCRGNRNRFVSQSDCLNFCIGEPV--VEPSAPITYSCAPEPEAGE 2162
QY 306 CTGPTSPHLVLMHYDPQRCGCMTPPARGCDGARGFETYEAGQAQACARGPG--DACVLPAY 364
Db 2163 CNRTT-----AMFYDESNACATAFTYTGCGGNGNRFETRDQCEBRCGFEKGVDCNEPYT 2218
QY 365 GQPCRGMEPRMAYSPLLQGCHPFVYGGCGNGNRFHRSRSCEDACBPVPTPCRCRLRS 424
Db 2219 TPCPTDMQTKYFYNTASQACEFTYGGCGDGTGNRFSDLEECOTVCLAGREPR----- 2270
QY 425 KLALSICRSDFAIVGRFTEVLEPEAAG 452
Db 2271 -----VGSKEICLLPVATG 2285

RESULT 3
044938 PRELIMINARY: PRT: 1572 AA.
AC 044938:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE THROMBOSPONDIN.
GN THRI.
OS Haemochus contortus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MOREDN;
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,
RA Knox D.P.;
RT "Cloning and characterization of thrombospondin, a novel multidomain
RT glycoprotein associated with the gut of Haemonchus contortus.";
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF043121; AAB99830.2; -.
DR HSSP: P05067; ITAM.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00014; Kunitz_BPTI; 6.
DR Pfam: PF00090; tsep_1; 6.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00131; KU; 6.
DR SMART: SM00209; TSP1; 7.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 6.
DR PROSITE: PS50092; TSP1; 2.
DR Serine protease inhibitor.
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2P903EC CRC64;

```

```

Query Match 8.6%; Score 259; DB 5; Length 1572;
Best Local Similarity 21.7%; Pred. No. 3,1e-12;
Matches 125; Conservative 51; Mismatches 184; Indels 216; Gaps 30;

QY 21 GLPLPGSHPGVCPNQLSP-NLWVDAQST-----CERECSRDQCAAE-KCCINVC 70
Db 990 GKTPARGSHKECPQGYTAYGCCPDGETALGRNDGCD-----DCYAYAGCCPD-- 1040
QY 71 GLHSCVAAFPSSPA-----APTTAASC-----EGFYCPQ-----QG 102
Db 1041 GESKAIQPDYACCPSYTLAPFLLGTVAPSKISSCALPDODGTVSSGGYKLWVYDTAG 1100
QY 103 SDCDMDGQVPC-----RCRDREKEPSFSCASDGLTYNNRCY----- 141
Db 1101 KCSQTFYGG--CDGNDNRFATEECCTITVEPRGT-----GRCLPYVEGRLCD 1148
QY 142 -----DAEACL-----RGLHLIYVCKHVLNMPSS-----PGPPTTARPPG 180
Db 1149 QPQAKYWDYNTKQCAAFWNRCHGN--ANNPASWECSFCKDVGPAIPTTAVP--- 1202
QY 181 AAPVPALYSFSPQAVQVGTASLH-----CDVSGRP-PAVTWEKQSHORE 227
Db 1203 -APQPPQIPHMQIESMEVEHQVVAHPIDEAPLKDNRNDLRPRPMPYIEEVCSTQDS 1261
QY 228 NLIM-RPDMYGNVYVTSIGOLVLYNA-----RPEDAGLYTCARH--AAGLLRADF 276
Db 1262 GPCODYSDFYDAYKGTG--QTFITYGGCGGNLNRFTEECQRCQCFELPTASSAYQA-- 1318
QY 277 PLVVOREPARDAAPSIAPAE-----CLPDVQACTGPTSP-- 312
Db 1319 -----QOSP LHGAQOEPRPHIQRGGAGAQSQOQEEYEQRPQPPYD-QHQGGVPRPP 1372
QY 313 -----HLVL-----WHYDQRCGCMTPPARGCDGAAR 339
Db 1373 PVVPOPSYGIHANSAKSHQCHLPLDVGKCGSFDPMYEMATGSCVEFKYSGCSGNAN 1432
QY 340 GFETYEACQQAACARGP-----GDACVLPAYGQPCRGMEPRMAYSPLLQGCHPFY 389
Db 1433 RFASRECECTVCRHSEPHSDTTSHTGSVCDEAKETGPTNATKWKYNKAGDTCRHFY 1492
QY 390 GCGCGNGNRFHRSRSC-----EDACPVPRT-PC 417
Db 1493 GCGCEGTRNRFNDQSCAKANHQDACTLPRYQGPC 1528

RESULT 4
076840 PRELIMINARY: PRT: 2167 AA.
AC 076840: 022911;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE C37C3.6 PROTEIN.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Felodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Giesel C., Bradshaw H.;
RT "The sequence of C. elegans cosmid C37C3.";
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
CC -1 ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC OF FORM B.
CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC -1 SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: U64857; AAC25868.1; -.
DR EMBL: U64857; AAC25867.1; -.

```

DR HSP, P00981; 1DTR.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR002223; Kunitz_BPTL.
 DR InterPro: IPR000884; TSP1.
 DR Pfam: PF00047; Ig_1.
 DR Pfam: PF00014; Kunitz_BPT1.
 DR Pfam: PF00090; TSP_1; 6.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00408; Ig_c2; 1.
 DR SMART: SM00131; KU; 11.
 DR SMART: SM00209; TSP1; 7.
 DR PROSITE: PS00280; BPTL_KUNITZ_1; 10.
 DR PROSITE: PS0279; BPTL_KUNITZ_2; 11.
 DR PROSITE: PS50092; TSP1; 4.
 DR Alternative splicing; Hypothetical protein: Serine protease inhibitor.
 KW VARSPLIC 147 155 MISSING (IN ISOFORM A).
 FT VARSPLIC 1556 1558 KDD -> SKF (IN ISOFORM A).
 FT VARSPLIC 1559 2167 MISSING (IN ISOFORM A).
 SQ SEQUENCE 2167 AA; 237599 MW; 96274786D52E3639 CRC64;

Query Match 8.3%; Score 251.5; DB 5; Length 2167;
 Best Local Similarity 21.6%; Pred. No. 1.7e-11;
 Matches 116; Conservative 48; Mismatches 190; Indels 183; Gaps 20;

QY 21 GLLPGSHRVCVCPNOLSPNLWVDAOSTGER--ECSRDDCAAEKCCINVCGLHSCVAA 78
 DB 1055 GELKAGPDGACCPPTTPPPFLMGTVAPHKIAACNQGES-----TYCGAGYKILAM 1107
 QY 79 RPPGSPAATTAAACEGFCPOGSDCDLWDGQPC-----RC-----RDRC 120
 DB 1108 HY-----DTGRGNQMYGGCGGNDNNFASQDMCETICEVPEPKGHCYLPKVDGHLRC 1161
 QY 121 EK-EPSFTCASDGLTYNR-----CYMDA-----EACRLGLHLHYPCKHVL 161
 DB 1162 DQLQPRY-----YDHSHKHCVAFWMRRGLGANNFNSFECC-----SMCKDVG 1206
 QY 162 SW-----PPSSGPPPTTAR--PTPGAAPVPALYSSPSQAVQ----- 198
 DB 1207 PYDAFTTAAPPPPOANAOYLPTPEVOIEIQSAEQPQPQOQOQOQOQOPRQS 1266
 QY 199 -----VGTGASL-----KHD-----VSG 211
 DB 1267 MEDICRSNQDAGPCETYSQDFYNAFSQECTFTYGGCGGLNFRSKDEEQRCEFHVG 1326
 QY 212 RPPAVVTWEK-----QSHQRENLMRPDMYGNVYVTSIGQLVLYNARPEDA----- 258
 DB 1327 AQPAAEQEAQAPAAPAPAPPSNIVSPPOQASAPVVPV-----NSKQDACHLNV 1379
 QY 259 -----GLYTCTARNAAGLLRADPPLSVQREPARDAAPSIPA 295
 DB 1380 DQGRKCAFDSWYEVATGSCVTFKRYGCGGANRFAASKQCCSLCYKPASEAASAGIDG 1439
 QY 296 PA--ECLPDVQACTGPTSPHLVLMHYDPORGCGWTFFPARCGDGAARGFETYEACQOACAR 353
 DB 1440 AAGINSYCDKADKDTGCPCTNFTVTKMYTKKADGTGCRNHFHYGGCGGTNNRFDWQCKKAC-Q 1498
 QY 354 GPGDACLPAVQPCRCRMEPRMAYSPLLQCHPPTVYGGCGEANGNNFHSRSECDACP 410
 DB 1499 NIKDACOLPRVQGPCSGKSHSYNTASHQCEFTFYGGCLGNTNRPATIEEQCARCP 1555

RESULT 5
 OGGORO
 ID OGGORO PRELIMINARY; PRT; 2174 AA.
 AC OGGORO;
 DT 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
 DE EXTRACELLULAR MATRIX PROTEIN PAPILIN PRECURSOR.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pteroyota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DP CN BM;
 RX PubMed-11076767;
 RA Kimerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
 RA Kimerov A.A., Kusche-Gulberg M., Kramer J.M., Ackley B.D.,
 RA Sieron A., Prockop D.J., Fessler J.H.,
 RT "Papilin in development: a pericellular protein with a homology to the
 RT ADAMTS metalloproteinases";
 RL Development 127:5475-5485(2000).
 CC -1 SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AF205357; AAC37995.1; -
 DR Flybase: FB900003137; Pn.
 DR InterPro: IPR003599; Ig_c2.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_1like.
 DR InterPro: IPR002223; Kunitz_BPT1.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00014; Kunitz_BPT1; 3.
 DR PRINTS: PR00003; ADISULPHORE.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00409; Ig_3.
 DR SMART: SM00408; Ig_c2; 2.
 DR SMART: SM00410; Ig_1like; 1.
 DR SMART: SM00131; KU; 3.
 DR SMART: SM00209; TSP1; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS50279; BPTL_KUNITZ_2; 3.
 DR PROSITE: PS50092; TSP1; 3.
 KW Matrix protein; Serine protease inhibitor; signal.
 FT SIGNAL 1 26 POTENTIAL.
 SQ SEQUENCE 2174 AA; 231936 MW; 038F07952623120 CRC64;

Query Match 8.3%; Score 251; DB 5; Length 2174;
 Best Local Similarity 24.7%; Pred. No. 1.9e-11;
 Matches 114; Conservative 32; Mismatches 179; Indels 136; Gaps 24;

QY 19 GAGLPLGLG-----SHRVCVCPNOLSPNLWVDAO--STCERE--CSRDDCA--AEK 64
 DB 1392 GCGLASPAVAKGAESENGCCPDGOTPASPNEGSGCGRERFGCCPSOTPARPNKEG 1451
 QY 65 CCINV--CGLHSCVAARPPGSPAAT--TASCEGFYCP-----QGSQDCDITWD 109
 DB 1452 CCIDTQFGCCPDNITLAARGPNNEGCECHYTPYC---CPDKSAATYNOEGCACE--T 1505
 QY 110 GQVPCRCRDR-----CEKEPSFTCASDGLTYNRCYMDAEACLRGLHLITVP 156
 DB 1506 TQYGC-CPKKTIAAGKPRNHEGCPCE--TFQFGCCPDGLTF-----AKGPHNHGCH 1552
 QY 157 CKNV-----LSWPPSP-GPPTTARPTGAAP--VPALYSSPSQAVQVGTASLH 206
 DB 1553 CTQTEFKCDDDEKTPAKPNEGCGTCEVSKFGCCPDGTAKTIDE-----KFGG----- 1600
 QY 207 GDVSGRPP-PAVTWKQSHQRENLMRP--DMTGNVYVTSIGQLVLYNARPE----- 257
 DB 1601 CENVQEPQKAGLEKTEGTGCTNNYSVKYFPTSGGCAREFYGGCGDNDNFESEARCKD 1660
 QY 258 -----AGLYCTARNAAGLLRADPPLSVQREPARDAAPSIPAECCLPDVQACTGPTSP 312
 DB 1661 TCQDVTGKHVCLLRKSAG-----PQTGFTKR 1686
 QY 313 HVLVLMHYDPORGCGWTFFPARCGDGAARGFETYEACQOACARGGG-DACVLPVAVQPCRGW 371
 DB 1687 -----WYFDVDRNRCEERFYGGCYGTNNRPFSLBQCGGTCAASENLPTCEBPVSGCAGN 1742
 QY 372 EPRMAYSPLLQCHPPTVYGGCGEANGNNFHSRSECDACPVP 412
 DB 1743 FERNYTDNEIDICRPTFYGGCGKKNKYPTHEACNYNCRCP 1783


```

RESULT 6
ID 045881 PRELIMINARY: PRT: 2225 AA.
AC 045881:
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE W01F3.3 PROTEIN.
GN W01F3.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN 1
RA SEQUENCE FROM N.A.
RP Cummings P.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN 12
RA SEQUENCE FROM N.A.
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: Z92815; CAB07294.1;
DR HSSP: P31713; 1SHP.
DR InterPro: IPR002899; EB.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00014; Kunitz_BPTI; 10.
DR Pfam: PF00086; thyroglobulin_1; 1.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00131; KU; 10.
DR SMART: SM00211; KU; 1.
DR SMART: SM00289; WRI; 4.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 8.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 10.
DR PROSITE: PS00484; THYROGLOBULIN_1; UNKNOWN_1.
KW Serine protease inhibitor.
SQ SEQUENCE 2225 AA; 242197 MW; A5DD8AE9D2A7B02A CRC64;

Query Match 7.8%; Score 236.5; DB: 5; Length 2225;
Best Local Similarity 20.0%; Pred. No. 2.9e-10;
Matches 123; Conservative 46; Mismatches 218; Indels 229; Gaps 28;

OY 28 SHPGVCPNQLSPNLVDAOSTGER---ECSDQDCAAEKCCINVC-----GLHSCVAA 78
DB 467 SLPGCHGHSFQRYFYNEISQCEQFTYSGCGNGNNGNYESREACEDRCAPPVGLPKCEI- 525
OY 79 RRPGRP-----AAPPTASCEGFVCPQOQSDCDINDGQVVC-----R 115
DB 536 ---GEPRLTKIGVPVNAKTD---CP-SGYRCSVVOHSSVCCPENKKVYGLOTSGARATR 578
OY 116 CRDCEKPEPSFTCASDGLTYV-----NRC-----YMDAE-ACLRGLH- 151
DB 579 CSLPRERGP---CDKYEELRFYNADLNCKRYFFMGCCGCEGNONNFERVECEBAC--GYVK 633
OY 152 -----LHIVCKHVLW-----PPSSP-GPEPTTARPPGAAP 183
DB 634 SGVTNRPMTEIRTTGQIRITPNGKLSWEETEDEEHAVPTTPIAPASAPTVAVSTQRAP 693

```

```

OY 184 VPPALXSSPSQAVVGGTASLHCDVSGRPAPVATWEXOSHORENLIMRPODMGVVYT 243
DB 694 VP-----TTPRPAPAVQTTTIRASRLTETTRVVKTEDEEEEEEV----- 736
OY 244 SIGOLVLYNARPEADGLYCTARNAGLIRADFPLSVQREPARDAASIPAPACL--- 300
DB 737 -----EEBEDG-----KEEPLVAVQPPVSQONTVLLGG 765
OY 301 -----PDVQACTGPTSP-----HLVLMHYDPQRCGCMTFPARGCDGAARGETYACQ 349
DB 766 IEDTTDSVNRCLHPRDSGNCRCGFVRWFEDDEKKKCDVFTYTCGCGGNNGNNSKEECWA 825
OY 350 ACARGP-----GDACVLPVAGPCRGMPERRAYSLLOQCHPFYVGGCEGNGNPH 400
DB 826 ICHKEPPTPSATPDPFSQVCSNDVDAGECNGVFERAFDAEADOCRAFTYGGCGGNGNNA 885
OY 401 SRESCEDACPVP-RPPPCAC-----RLRSKLALSLCRS-----D 434
DB 886 TWQECRSRCVAMAKKSPVATCADIEVGECAVFSRFAPDRKINNACRSFTYGGCGGNGN 945
OY 435 FAIVGRLE-----VLEEPFAAGGIARVALBDV-----LKDDKMLKELGTLYLEVTL 484
DB 946 FATLQECTKCKVNRGVCPEPPAC-----DTNRCOLVNDRSCPF----- 984
OY 485 KMWACPCPRMNTAGDGP 500
DB 985 ---CSCPVPKQASPP 996

RESULT 7
ID 017644 PRELIMINARY: PRT: 1043 AA.
AC 017644:
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE C34F6.1 PROTEIN.
GN C34F6.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN 1
RA SEQUENCE FROM N.A.
RA White S.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN 12
RA SEQUENCE FROM N.A.
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: Z81479; CAB03944.1;
DR HSSP: P10646; 1TXX.
DR InterPro: IPR002899; EB.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 8.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00131; KU; 8.
DR SMART: SM00289; WRI; 6.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 8.

```

KW Serine protease inhibitor.
 SQ SEQUENCE 1043 AA; 115537 MW; F25F5EB7FA1F4A5 CRC64;

Query Match 7.2%; Score 218; DB 5; Length 1043;

Best Local Similarity 20.9%; Pred. No. 3.8e-09;

Matches 100; Conservative 46; Mismatches 155; Indels 178; Gaps 20;

```

OY 53 CSRDODCAAEKCCINVCGLHSCVAFPGSPAPPTAASCEGF----- 96
DB 638 CSDPERCPSTHFCHIGVEGSENYCKPH-GDPCQALAGTSGFSTIRYKETRRCDD 696
OY 97 -----VCPQSGSDCDLMDGQVPCRCRDCEKEPSFTCASDGLTY 135
DB 697 FSTEGGKGNANNNFLLEDGLVCPVLPNPSCL--GEPLT-----SLQKEP-VICGGE---- 745
OY 136 YNRCYMDAECALRGHLHIVPCKHVLWSPSSPGPETTARPTGAAPVPPALYSSPSQ 195
DB 746 -----DTCRNGYICHV-----GGAPEET-NCCPGT-----RRPCDL 775
OY 196 AVQVG-GTASL-----HCDVSGRPPPAVTWKEKSHORENLIMRPDQMYG 238
DB 776 PLEVGGVGEKLEKRMFPGDGIQMCPRFVYKMGKGNNNFLT--KQSCROSKEMNFCG-YG 832
OY 239 NVVYVSIQGLV-----YNAREPAGLYTCTARNAAGLLRADPLSVYQREPARDAAPS 292
DB 833 DPLVDTGGRMLCTGGQRYNSCPONS--YCHVSSALTTLTLC-----PKRKIDP- 879
OY 293 IPAPAECLPDVQACTGPTSPHLVLMHYDPORGCMTFPARGDDGAAREFETYEACQA- 350
DB 880 -----CQAVVEGTS--SEDLPRWFPRKONRCAPFTYGGVAGENNENFISQNTCMGCP 931
OY 351 -----CAKPGDACVLPAVQGP 367
DB 932 EYRNYCPHGIPLEGSTVTSQIDKGCPEGFTCHMSSEFNYSICQDPMDCLTSARDGP 991
OY 368 CRGMEPRMAYSPILQOCHPEFYVGGCEGNGNFSRESEDCPPRPTRPCACLRSKL 426
DB 992 CNNEFKRQYDANTDTCVEYQYGGCEGTLNPHSLDCTETC-----CKEYRRHRL 1043

RESULT 8
OY 09QW87 PRELIMINARY; PRT; 144 AA.
AC 09QW87;
DB 01-MAY-2000 (TREMblrel. 13, Created)
DB 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DB 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DB BIKUNIN.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE.
RX MEDLINE=97018241; PubMed=8864857;
RA Yamamoto T., Yamamoto K., Sinohara H.;
RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian
RT hamster urine and plasma.";
RT J. Biochem. 120:145-152(1996).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DB HSSP: P02760; IBIK.
DB InterPro: IPR002223; Kunitz_BPTI.
DB Pfam: PF00014; Kunitz_BPTI. 2.
DB PRINTS: PR00759; BASICTPASE.
DB SMART: SM00131; KU: 2.
DB PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DB PROSITE: PS0279; BPTI_KUNITZ_2; 2.
KW Serine protease inhibitor.
SQ SEQUENCE 144 AA; 15700 MW; C71661C10FC2AC5 CRC64;
```

Query Match 7.2%; Score 216.5; DB 11; Length 144;

Best Local Similarity 35.5%; Pred. No. 5.7e-10;

Matches 44; Conservative 16; Mismatches 57; Indels 7; Gaps 2;

```

OY 294 PAPACLPDVQAC-----TGPTSPHLVLMHYDPORGCMTFPARGCGAAREFETACQ 348
DB 14 PLYVDVLRKESCOLSISYSGPLCMTERKYNNGASMACETFFHYGGCLGNGNFSRECECL 73
OY 349 QACARPGDACVLPAVQGPCRGMEPRMAYSPILQOCHPEFYVGGCEGNGNFSRECECA 408
DB 74 QTCRTVA--ACSLPIVQGPCRAYVELMAFDAQCKCYQFYGGCKGNGNFTSECKEY 131
OY 409 CPVP 412
DB 132 CGVP 135

RESULT 9
OY 021418 PRELIMINARY; PRT; 922 AA.
AC 021418;
DB 01-NOV-1996 (TREMblrel. 01, Created)
DB 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DB 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DB K10D3.4 PROTEIN.
GN K10D3.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.;
RU Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Beirs M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latteille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RT Nature 368:32-38(1994).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DB EMBL: Z75545; CAA9886.1; -.
DB HSSP: P10646; IPFX.
DB InterPro: IPR002899; EB.
DB InterPro: IPR002223; Kunitz_BPTI.
DB Pfam: PF01683; EB; 1.
DB Pfam: PF00014; Kunitz_BPTI; 5.
DB PRINTS: PR00759; BASICTPASE.
DB SMART: SM00131; KU: 5.
DB SMART: SM00289; WR1; 6.
DB PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DB PROSITE: PS0279; BPTI_KUNITZ_2; 5.
KW Serine protease inhibitor.
SQ SEQUENCE 922 AA; 99995 MW; 0CCBCC0AE2524CFD CRC64;
```

Query Match 7.2%; Score 216; DB 5; Length 922;

Best Local Similarity 19.7%; Pred. No. 4.9e-09;

Matches 109; Conservative 45; Mismatches 166; Indels 234; Gaps 24;

```

OY 31 GVCPE-NQISPLMWD-----AOSTCERCSRDQCAAAKCCINVCGLHSCVAAAF 80
DB 109 GFCEPLNNGNPRFRIENVOOCYGGADATCAIGALSTDC-----ICDSDCTVNN- 158
```

```

OY 81 PGSPAFTTASG-----EGFYCPQGSDDCDIMDGPV---CR-----CR 117
DB 159 PLSFCCPSRAAFACIQIPNEGYYTPGGGTTLMHWHDPITGECREKYOGYGANNFQTK 218
OY 118 DRCEKEPFTSCADGLTFYNN-----RCYMDAACLRLGLHLHIVPCKHVLSPPS 166
DB 219 DHCESEYCNQTC-NRGLPLXRDRTGVKQEPVYCCGNDNGC-----N 258
OY 167 SPGPETTARPTGAP-----VPPALYSPPQAVQVGGTSLHCDV-----S 210
DB 259 NPNYOCTMTGLQCCPTFLFICSRNGGIPSEVYNT-----AGLPTVEFDVGIPIPGS 311
OY 211 GRRPAVWEMKOSHORENLMRPDMYGNVVTISGLVLYNAREDAGLYCTIRNAG 270
DB 312 GTTSREYDYRE-----GRCIOFS-----YIGOGNERN 341
OY 271 LLRADFPPLSVQR-----EPARDA-----APSIPAPAC-----299
DB 342 FLSDHCKEFCRILCSAGEPLKDSGGERNMECSPTSGGANSCEPSTHSCSTSGSTTFGG 401
OY 300 -----LPDVQACTGPTSPHLVLMHIDPQRGCGMTEPARCGDGAARGFETYEACQ 348
DB 402 VCCPRPQYVCKLPREGNGCTYSNR---WMFNAKTGNCSEEFYSGCGGNANFETYEKCCQ 458
OY 349 QAC--ARG-----PGDA-- 358
DB 459 DYCRRARSEPQICGTALTDSNGNFIICGSSAATSTCPRANHYYDGTYYGCCPTQAVT 518
OY 359 CVLPAPVOC-PCRGEPRNAYSPLLQOCHPFYVGGCEGNNGNHFRESCEDACPVPRTP-- 415
DB 519 CLSYKSGASGSPAYTRWYDSTRTTCQYISYNGCDGNSNNRATQDDCKDCRVASCPDG 578
OY 416 -----PCRAC 420
DB 579 GEVWKEONGAARAC 592

RESULT 10
P78491 PRELIMINARY; PRT; 151 AA.
AC P78491:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALHA-1-MICROGLOBULIN (FRAGMENT).
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91214554; PubMed=1708673;
RA Veer H., Gebhard W.;
RT "Structure of the human alpha 1-microglobulin-bikunin gene.";
RL Biol. Chem. Hoppe-Seyler 371.1186(1990).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; X54818; CA38587.1; -.
DR HSSP; P02760; 1BTK.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI.2.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 2.
KW Serine protease inhibitor.
FT NON_TER 1
SQ SEQUENCE 151 AA; 16542 MW; 88F400C5ECA19831 CRC64;

```

```

Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;
OY 306 CTGPTSPHLVLMHIDPQRGCGMTEPARCGDGAARGFETYEACQACARGDQACVLPVQ 365
DB 39 CMGMSRYF-----YNGTSMACETFGYGGCMGNGNFFVEKEDELQCTRTVA--ACNLPIVR 92
OY 366 GPCRMEPRNAYSPLLQOCHPFYVGGCEGNNGNHFRESCEDACVP 412
DB 93 GPCRATQLMAFDVAVKGVLPFYGGCGNGKNGFTSEKREYCGVP 139

RESULT 11
O9DBJ9
ID O9DBJ9 PRELIMINARY; PRT; 349 AA.
AC O9DBJ9:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALHA 1 MICROGLOBULIN/BIKUNIN.
DE AMP.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=LIVER.
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Akawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Salto T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Salto R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Winking L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlski S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AK004907; BAB3659.1; -.
DR MGI; MGI:88002; Ampd.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000566; Lipocalin_cytfabp.
DR Pfam; PF00014; Kunitz_BPTI.2.
DR Pfam; PF00061; lipocalin.1.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR00179; LIPOCALIN.
DR SMART; SM00131; KU; 2.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Serine protease inhibitor.
SQ SEQUENCE 349 AA; 39101 MW; E593222FBC01BC3D CRC64;

Query Match 6.7%; Score 203; DB 11; Length 349;
Best Local Similarity 40.6%; Pred. No. 1.8e-08;
Matches 39; Conservative 13; Mismatches 42; Indels 2; Gaps 1;
OY 317 WHYDPQRGCGMTEPARCGDGAARGFETYEACQACARGDQACVLPVQPCRGMEPRNA 376
DB 246 YYNAGSMACETFGYGGCGNGNFFSEKDELQCTRTVA--ACNLPIVQGPCRAPIKLMA 303

```

QY 377 YSEPLQOCHPEYVYGCCEGNGNNHSHSCEDACVPV 412
 Db 304 FDAAGCKCIOFYHGGCKNGNKFYSKECKEYCGVP 339

RESULT 12

Q9XWXS PRELIMINARY: PRT: 1743 AA.

AC Q9XWXS; PRELIMINARY: PRT: 1743 AA.
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Y438B.3 PROTEIN.
 GN Y438B.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN RP
 RA SEQUENCE FROM N.A.
 RA Ainscough R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierrey-Wieg J., Thomas K., Vaudin M., Vauden K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spratt K., Wohlman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: A0032623; CAZ21511.1; .
 DR HSSP: P03067; 17AW.
 DR Interpro: IPR002899; EB.
 DR Interpro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 16.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 15.
 DR SMART: SM00289; KRL; 10.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 15.
 DR Serine protease inhibitor.
 KW S04
 SQ SEQUENCE 1743 AA; 19111 MW; B93C163556433C2A CRC64;

Query Match 6.7%; Score 202.5; DB 5; Length 1743;
 Best Local Similarity 22.8%; Pred. No. 1.2e-07;
 Matches 124; Conservative 49; Mismatches 235; Indels 153; Gaps 26;

QY 52 ECSRDDPCAAREKCIIVWGLHSCVAFRFGSPAAP-----TTAASCE 94
 Db 1019 QCSPOQVCPGSHCHVTEANYCKA--LGGDPCGQDLRGVGSQLSRRYMQOSQCC 1076
 QY 95 GF-VCPQGG-----DCDIWDGQFVCRCDRCERKESFTCASDGLTYNR---CYMD 142
 Db 1077 PFSYCGCKGTQNNFLTKQDCD-----RTCYELDNCAALGDPMQNNRPLOCSNT 1126
 QY 143 AEACLGILHILYPCCKHYLSWPPSSGPPETARPT---PGAAPVPALYSSESPOAVOV 199
 Db 1127 ASTCGAQFMCHEGANDDTYVCCGGRVESQICQAPAVGTGATLPRWYVNAQTMCQVQF 1186
 QY 200 GGTASLHCDVSGRPVATWEEKSHORENLIIMPDOMYGNVYVTSIGOLVILNAREDDG 259
 Db 1187 -----VYAGR-----GNQNNRFSQACQCEQTCFYVYVNCPTGSPMLDASTNKP----- 1229

QY 260 LYTCT-ARNAGLLRADE--PLSVQRE-PARDAPSPAPAECLPDVOACTGPTSPHLV 315
 Db 1230 -VPTFGNSCG---ADWCHLGLVPDEYQCCPSGPTNPAGCQGLPSEGVGTGAPAPTS 1285
 QY 316 LMHYDPORGCMTPPARCGDGAARGFETYEACQAC-----ARGGDA 358
 Db 1286 RMYDDQTMCKQFTYNGRNGNQNFLQEDCATCTCVTFNPQNPITALPATLCSGSS 1345
 QY 359 -----CVLPAYQGR---CRGWEPKRWANSPLLQCHPEYVYGCCEGNGNNHSHSCED 407
 Db 1346 DTGCANMWHCHIGANQDSTVCC---PSGRKSKNFQCCLPQYMKRNGNQNFEQACER 1401
 QY 408 ACVPRTPT-----PCRACRLRSKILSLCRS--DRAIVGRLE----- 443
 Db 1402 TCFVFINPCLGEVILEGAKRPCK-----PLKNSGSGSTFCHGNPSDQNSFCCPR 1455
 QY 444 VLEED-----BAAGIARVALEDVLDK-----DKMGLK-----FLGTRELVLTSGMD 486
 Db 1456 INDDPCNAFVNRNGENFMTRYNNPVEGDGFQYRGLKGNENNFLLTKMCECTCKPIL 1515
 QY 487 WACPCPMWTAGDGPLVIMGEV 507
 Db 1516 TAC-----FGGESPLMNGRY 1531

RESULT 13

Q70160 PRELIMINARY: PRT: 352 AA.

AC Q70160; PRELIMINARY: PRT: 352 AA.
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ALPHA-1-MICROGLOBULIN/BIKUNIN.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 OX NCBI_Taxid=10141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=LIVER;
 RA Yoshida K., Suzuki Y., Yamamoto K., Sinohara H.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AB006444; BAA25305.1; .
 DR HSSP: P02760; 1BIR.
 DR Interpro: IPR001064; Crystallin.
 DR Interpro: IPR002223; Kunitz_BPTI.
 DR Interpro: IPR002345; Lipocalin.
 DR Interpro: IPR000566; Lipocalin_cytfabp.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR Pfam: PF00061; Lipocalin; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR PRINTS: PR00179; LIPOCALIN.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 2.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 KW Serine protease inhibitor.
 SQ SEQUENCE 352 AA; 39033 MW; BAA7447B574C2FAB CRC64;

Query Match 6.5%; Score 196; DB 11; Length 352;
 Best Local Similarity 36.2%; Pred. No. 6.8e-08;
 Matches 38; Conservative 13; Mismatches 52; Indels 2; Gaps 1;

QY 308 GPTSHLVLMYHMDPQRGCMFPARGCDGAARGFETYEACQACARGDGACVLPAYQGP 367
 Db 238 GPCGLGLFRIRYFNGSSMACEIFHYGCGLGNNGNNFSEKCELOTCTRYA--ACNLPYVGP 295
 QY 368 CRGWEPKRWANSPLLQCHPEYVYGCCEGNGNNHSHSCEDACVPV 412
 Db 296 CGGSAQLMAFADAIKRCVRFYTGCGGNGNKFYSKECKEYCGAP 340

RESULT 14
ID 061893 PRELIMINARY: PRT: 1965 AA.
AC 061893;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE T07H8.4 PROTEIN.
GN T07H8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Smaiden N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Gattung S., Megal L.;
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Waterston R.;
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF069945; AAC17681.1; .
DR HSPD: P00981; IDTK.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 11.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00131; KU; 14.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 5.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 14.
DR PROSITE: PS01186; EGF_2; UNKNOWN_2.
DR PROSITE: PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat;
KW Serine protease inhibitor.
SQ SEQUENCE 1965 AA; 223504 MW; 80E22ACB1CE52234 CRC64;

Query Match 6.5%; Score 195; DB 5; Length 1965;
Best Local Similarity 32.8%; Pred. No. 5.5e-07;
Matches 43; Conservative 23; Mismatches 61; Indels 4; Gaps 3;

QY 316 LMHYDPOBGCMTFPAAGCDGAARGFETYEACQACARPGACVLPVAVG- CRGWEPR 374
DB 1748 VFVYTLTRGTCDQPLVGGCGGNPNRPFDTICQACAEVTGTDCMESLDGSGWCEAMSNR 1807
QY 375 WAVSPDLQCHPFFVYGCGEGNGNNEFHSRSCEDAC-VPRTPPC-RACRLSKLALSLC 431
DB 1808 YFENRARRCKGPFHYTGCGCKSGNNPLTKRECQTKCEKRPRAAPSKKAKLAKIIFIIFEO 1867

QY 432 RSDFAIVGRLT 442
DB 1868 NSRYNVTGRDT 1878
RESULT 15
ID 093424 PRELIMINARY: PRT: 287 AA.
AC 093424;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 33.1 KDA PROTEIN.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Gracey A.Y.;
RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF008648; AAC19410.1; .
DR HSPD: P31713; ISHP.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 3.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00131; KU; 3.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 3.
KW Hypothetical protein; Serine protease inhibitor.
SQ SEQUENCE 287 AA; 33093 MW; DF69B3D76718115E CRC64;

Query Match 6.4%; Score 194.5; DB 13; Length 287;
Best Local Similarity 35.9%; Pred. No. 7.1e-08;
Matches 37; Conservative 14; Mismatches 51; Indels 1; Gaps 1;

QY 308 GPTSPILVLMHYDPOBGCMTFPAAGCDGAARGFETYEACQAC-ARGGDACVLPVAVG 366
DB 49 GPCKALKDRFVFDPTDGRCESEFYGCGGNGNNEFETLQCEKCKLVKEDKSPQQLDDEPG 108
QY 367 PCRGWEPRWAVSPLDQCHPFFVYGCGEGNGNNEFHSRSCEDAC 409
DB 109 PCRGVLPRTFFPKSQCEKRFYGGCGFANNFXTIKECHERC 151

Search completed: February 26, 2002, 01:20:09
Job time: 599 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:10:20 ; Search time 67.78 Seconds

(without alignments)
296.434 Million cell updates/sec

Title: US-09-819-136-2

Sequence: 1 MPALRPLPLLLLRLLRSGA.....KRIELLEKQACELNRFQD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	7.3	346	1	AMBP_MERUN
2	218.5	7.2	349	1	AMBP_MESAV
3	214	7.1	337	1	AMBP_PIG
4	214	7.1	349	1	AMBP_RAT
5	213	7.1	352	1	AMBP_HUMAN
6	210	7.0	352	1	AMBP_BOVIN
7	208	6.9	123	1	IATR_HORSE
8	205	6.8	123	1	IATR_SHEEP
9	204	6.8	300	1	TEPI_RABIT
10	203	6.7	349	1	AMBP_MOUSE
11	194.5	6.4	304	1	TEPI_HUMAN
12	192	6.4	252	1	SPR2_HUMAN
13	192	6.4	252	1	SPR2_MOUSE
14	188	6.2	1416	1	YB1_CAEEL
15	186.5	6.2	306	1	TEPI_MOUSE
16	184.5	6.1	302	1	TEPI_RAT
17	179.5	6.0	235	1	TEPI_HUMAN
18	179.5	6.0	235	1	TEPI_MACMO
19	179	5.9	1959	1	AGRI_RAT
20	176.5	5.9	2944	1	CA17_HUMAN
21	176	5.8	110	1	IBP_CARCR
22	173.5	5.8	513	1	SPR1_HUMAN
23	167.5	5.6	62	1	IP52_ANESU
24	166.5	5.5	355	1	AMBP_PLEPL
25	164	5.4	1955	1	AGRI_CHICK
26	163	5.4	2481	1	UN52_CAEEL
27	162.5	5.4	591	1	GRN_CAVPO
28	161	5.3	230	1	TEPI_MOUSE
29	158	5.2	61	1	IBP1_TACTR
30	158	5.2	507	1	SPR1_MOUSE
31	158	5.2	1051	1	PRK7_CHICK
32	158	5.2	1070	1	PRK7_HUMAN
33	157.5	5.2	3176	1	CA36_HUMAN

34	157.5	5.2	4393	1	PCBM_HUMAN	P98160 homo sapien
35	157	5.2	2318	1	NTC3_MOUSE	O61982 mus musculus
36	155	5.1	69	1	CRPT_BOOMI	P81162 boophilus m
37	154	5.1	898	1	FAS2_SCHAM	P22648 schistocerc
38	152.5	5.1	1906	1	KMUS_CHICK	P11799 gallus gall
39	152	5.0	1328	1	AGRI_DISOM	O90404 discopyge o
40	151	5.0	67	1	IBPC_BOVIN	P00976 bos taurus
41	151	5.0	133	1	EPPI_HUMAN	O95925 homo sapien
42	151	5.0	944	1	ROB2_MOUSE	O92138 mus musculus
43	150	5.0	134	1	EPPI_MOUSE	O94601 mus musculus
44	149	4.9	63	1	IMAP_DROFU	P1424 drosophila
45	148	4.9	3707	1	PCBM_MOUSE	O05793 mus musculus

ALIGNMENTS

RESULT	ID	AMBP_MERUN	STANDARD:	PRT:	346 AA.
AC	062577	062576			
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	AMBP	PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].			
GN	AMBP	OR ITTL.			
OS	Meriones unguiculatus (Mongolian jird).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Gerbillinae;				
OC	Meriones.				
OX	NCBI_TaxID=10047;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=95110820; PubMed=7529051;				
RA	Ide H., Itoh H., Nawa Y.;				
RT	"Sequencing of cDNAs encoding alpha 1-microglobulin/Dikunin of Mongolian gerbil and Syrian golden hamster in comparison with man and other species."				
RL	Biochim. Biophys. Acta 1209:286-292(1994).				
CC	- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA AND ALBUMIN (BY SIMILARITY).				
CC	- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTTIC ELASTASE (BY SIMILARITY).				
CC	- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO SEPARATELY FUNCTIONING PROTEINS.				
CC	- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY SIMILARITY).				
CC	- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN FAMILY.				
CC	- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL: D31813; BAA0600.1; -				
DR	HSSP: P10646; IAD2.				
DR	InterPro: IPR002223; Kunitz_BPTI.				
DR	InterPro: IPR002345; Lipocalin.				
DR	InterPro: IPR000566; Lipocalin_cytrFAB.				
DR	Pfam: PF00014; Kunitz_BPTI; 2.				
DR	Pfam: PF00061; Lipocalin; 1.				
DR	PRINTS: PR00179; LIPOCALIN.				
DR	PRINTS: PR00759; BASICPTASE.				

DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 Lipocalin.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 202 ALPHA-1 MICROGLOBULIN.
 FT CHAIN 205 346 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
 CHAIN.
 FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 FT BINDING 286 336 BPTI/KUNITZ INHIBITOR 2.
 FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 FT DISULFID 90 187 BY SIMILARITY.
 FT DISULFID 230 280 BY SIMILARITY.
 FT DISULFID 239 263 BY SIMILARITY.
 FT DISULFID 255 276 BY SIMILARITY.
 FT DISULFID 286 336 BY SIMILARITY.
 FT DISULFID 295 319 BY SIMILARITY.
 FT DISULFID 311 332 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 INHIBITORY SITE (PI) (CHYMOTRYPSIN,
 ELASTASE) (BY SIMILARITY).
 FT ACT_SITE 240 241 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT SEQUENCE 346 AA; 38643 MW; FLA463810918D5F CRC64;
 SO
 Query Match 7.3%; Score 219; DB 1; Length 346;
 Best Local Similarity 40.0%; Pred. No. 3.4e-07;
 Matches 42; Conservative 13; Mismatches 48; Indels 2; Gaps 1;
 QY 308 GPTSPHLVLMHYDPORGCGMTFPAAGCDGARGFEYACQACARGPDCVLPAYOGP 367
 DB 237 GPCGAMERHYNGTSMACETFOYGGCGNGNNFISEKCLQTCRTVA--ACNLPYOGP 294
 QY 368 CRGEPKRWASPLQOCHPEYVGGCEGNGNNFHSRESCDACPVP 412
 DB 295 CRAVTKLWADPAQGCIOFTYGGCKGNKFESEKEKCYGVP 339
 RESULT 2
 AMBP_MESAU STANDARD; PRT; 349 AA.
 AC 060559; 060558;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 TRYPsin INHIBITOR LIGHT CHAIN (ITI-1C) (BIKUNIN) (HI-30)].
 GN AMBP OR ITIL.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_Taxid=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95110820; PubMed=7529051;
 RA Ide H., Itoh H., Nawa Y.,
 RT "Sequencing of cdnas encoding alpha 1-microglobulin/bikunin of
 RT Mongolian gerbil and syrian golden hamster in comparison with man and
 RT other species";
 RL Biochim. Biophys. Acta 1209:286-292(1994).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN (BY SIMILARITY).
 CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE (BY SIMILARITY).
 CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO

CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D31814; BA06601.1; -
 CC HSSP; P10646; ITEX.
 CC InterPro; IPR002223; Kunitz_BPTI.
 CC InterPro; IPR002345; Lipocalin.
 CC InterPro; IPR000566; Lipocalin_cytrpab.
 CC Pfam; PF00014; Kunitz_BPTI; 2.
 CC Pfam; PF00061; Lipocalin; 1.
 CC PRINTS; PR00179; Lipocalin.
 CC PRINTS; PR00759; BASICPTASE.
 CC SMART; SM00131; KU; 2.
 CC PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 CC PROSITE; PS00279; BPTI_KUNITZ_2; 2.
 CC PROSITE; PS00213; LIPOCALIN; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 KW Lipocalin.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
 FT CHAIN 205 349 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
 CHAIN.
 FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 FT BINDING 286 336 BPTI/KUNITZ INHIBITOR 2.
 FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 FT DISULFID 90 187 BY SIMILARITY.
 FT DISULFID 230 280 BY SIMILARITY.
 FT DISULFID 239 263 BY SIMILARITY.
 FT DISULFID 255 276 BY SIMILARITY.
 FT DISULFID 286 336 BY SIMILARITY.
 FT DISULFID 295 319 BY SIMILARITY.
 FT DISULFID 311 332 BY SIMILARITY.
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT ACT_SITE 240 241 INHIBITORY SITE (PI) (CHYMOTRYPSIN,
 ELASTASE) (BY SIMILARITY).
 FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT SEQUENCE 349 AA; 38782 MW; 8C954584B7DBE728 CRC64;
 SO
 Query Match 7.2%; Score 218.5; DB 1; Length 349;
 Best Local Similarity 32.9%; Pred. No. 3.7e-07;
 Matches 50; Conservative 24; Mismatches 67; Indels 11; Gaps 4;
 QY 266 RNAAGLIRADPFLSVOREPARDAAPSIAPACLPDVQAC-----TGPTSPHLVLMHYD 320
 DB 194 QKSTSLRA--RRAVLPQE--NEGSGTGLVTDVLKREDSQLSYSEGPCLGIEMERYYN 249
 QY 321 PORGCGMTFPAAGCDGARGFEYACQACARGPDCVLPAYOGPCKGMEPRWASPL 380
 DB 250 GASMACETFFYGGCLGNGNNFSEKCLQTCRTVA--ACSLPIYOGPCRAYVELMAFDAA 307
 QY 381 LQOCHPEYVGGCEGNGNNFHSRESCDACPVP 412
 DB 308 QGKCVQFSTYGGCKGNKFESEKEKCYGVP 339
 RESULT 3
 AMBP_PIG STANDARD; PRT; 337 AA.
 ID

AC P04366; P34954;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30) (EI-14)]
 DE (FRAGMENT).
 GN AMP OR ITIL.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90353595; PubMed=1696914;
 RA Gebhard W., Schreimüller T., Vetr H., Wachter E., Hochstrasser K.;
 RT "Complementary DNA and deduced amino acid sequences of porcine alpha
 RT 1-microglobulin and bikunin."
 RL FEBS Lett. 269:32-36(1990).
 RN [2]
 RP SEQUENCE OF 2-337 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=91113729; PubMed=1703444;
 RA Tavaakkol A.;
 RT "Molecular cloning of porcine alpha 1-microglobulin/HI-30 reveals
 RT developmental and tissue-specific expression of two variant messenger
 RT ribonucleic acids."
 RL Biochim. Biophys. Acta 1088:47-56(1991).
 RN [3]
 RP SEQUENCE OF 212-334.
 RX MEDLINE=8525967; PubMed=2408637;
 RA Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;
 RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
 RT the inter-alpha-trypsin inhibitor, X. The amino-acid sequences of the
 RT trypsin-released inhibitors from horse and pig inter-alpha-trypsin
 RT inhibitors."
 RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN.
 CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE.
 CC -1- PFM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PFM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X53685; CAA37725.1; -;
 DR EMBL; X52087; CAA36306.1; -;
 DR PIR; A01208; TIRGBI.
 DR PIR; S1066; S11066.
 DR HSP; P10646; ITRX.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR000566; Lipocalin_cytFABP.
 DR Pfam; PF00014; Kunitz_BPTI_2.
 DR Pfam; PF00061; lipocalin_1.
 DR SMART; SM00131; KU_2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;

KW Lipocalin.
 FT NON_TER 1 1
 FT SIGNAL <1 188
 FT CHAIN 5
 FT CHAIN 191 337
 FT DOMAIN 216 266
 FT BINDING 272 322
 FT BINDING 38 38
 FT DISULFID 76 173
 FT DISULFID 216 266
 FT DISULFID 225 249
 FT DISULFID 241 262
 FT DISULFID 272 322
 FT DISULFID 281 305
 FT DISULFID 297 318
 FT CARBOHYD 100 100
 FT CARBOHYD 235 235
 FT ACT_SITE 226 227
 FT ACT_SITE 282 283
 FT CONFLICT 49 49
 FT CONFLICT 259 259
 FT CONFLICT 270 270
 FT CONFLICT 278 278
 FT CONFLICT 283 283
 FT CONFLICT 285 286
 FT CONFLICT 293 293
 FT CONFLICT 311 311
 FT CONFLICT 315 315
 SQ SEQUENCE 337 AA; 37690 MM; 1F630FF98B3C070F CRC64;
 Query Match 7.1%; Score 214; DB 1; Length 337;
 Best Local Similarity 39.0%; Pred. No. 6.9e-07;
 Matches 41; Conservative 15; Mismatches 47; Indels 2; Gaps 1;
 QY 308 GPTSHLVIMHNDPPRGCGMTFPAAGCGAAGFEYENCCQACARGPDACVAPVQGP 367
 DB 223 GGLCMKIRKRYFYNGSGSMACETFRYGGCMGNFNFSSEKCLDTCR--TVEACSLIVSGP 280
 QY 368 CGWGPFRMAYSPDLOCHPFYVYGCCEGNGNNFHSRSCEDACPVP 412
 DB 281 CRGFFQLMAFDVAGKCVLFNNGCGGNGNGYSEKECKEYGV 325
 RESULT 4
 AMBP_RAT
 ID AMBP_RAT STANDARD; PRT; 349 AA.
 AC 064240; 063336;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
 GN AMP OR ITIL.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92182014; PubMed=1371936;
 RA Lindqvist A., Bratt T., Altiert M., Kasteren W., Aakerstrom B.;
 RT "Rat alpha 1-microglobulin: co-expression in liver with the light
 RT chain of inter-alpha-trypsin inhibitor."
 RL Biochim. Biophys. Acta 1130:63-67(1992).
 RN [2]
 RP SEQUENCE OF 141-195 FROM N.A.
 RX MEDLINE=87033744; PubMed=2429963;
 RA Kasteren W., Bioerck L., Aakerstrom B.;
 RT "Developmental and tissue-specific expression of alpha 1-microglobulin

```
RT      mRNA in the rat." ;  
CC CC -I- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL  
CC CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPIINAL FLUID. IT  
CC CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA  
CC CC AND ALBUMIN (BY SIMILARITY).  
CC CC -I- FUNCTION: INNER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND  
CC CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCTYTIC  
CC CC ELASTASE (BY SIMILARITY).  
CC CC -I- PTM: THE PRECURSOR IS PROTEOLYTIALLY PROCESSED INTO TWO  
CC CC SEPARATELY FUNCTIONING PROTEINS.  
CC CC -I- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY  
CC CC SIMILARITY) .  
CC CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN  
CC CC FAMILY.  
CC CC -----  
CC CC -I- SIMILARITY: CONTAINS 2 BPIT/KUNITZ INHIBITOR DOMAINS.  
CC CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outposton -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (see http://www.isb.ch/announce/  
CC CC or send an email to license@isb-sib.ch).  
CC CC -----  
DR DR EMBL, S87544; AAA21782.1; -.  
DR DR EMBL, J02600; AAA41596.1; -.  
DR DR HSSP; P12111; IKUN.  
DR DR InterPro: IPR002223; Kunitz_BPIT.  
DR DR InterPro: IPR002345; Lipocalin.  
DR DR InterPro: IPR000565; Lipoclin.CytfABP.  
DR DR Pfam; PF00014; Kunitz_BPIT_2.  
DR Dr Pfam; PF00061; lipocalin; 1.  
DR DR PRINTS; PR00179; LIPOCALIN.  
DR DR PRINTS; PR00759; BASICPTASE.  
DR DR SMART; SM00131; KU; 2.  
DR DR PROSITE; PS00280; BPIT_KUNITZ_1; 2.  
DR DR PROSITE; PS00279; BPIT_KUNITZ_2; 2.  
DR DR PROSITE; PS00213; LIPOCALIN; 1.  
KW Glycophoritin; Plasma; Signal; Serine protease inhibitor; Repeat;  
KW Lipocalin.  
FT FT SIGNAL                     1          19  
FT FT CHAIN                    20         202  
FT FT CHAIN                   205        349  
  
            BY SIMILARITY.  
            ALPHA-1-MICROGLOBULIN.  
            INTER-ALPHA-TRYPSIN INHIBITOR LIGHT  
            CHAIN.  
FT FT DOMAIN                 230        280  
FT FT BINDING                286        336  
FT FT     52           52  
FT FT DIISUFID               90       187  
FT FT DIISUFID              230       280  
FT FT DIISUFID              239       263  
FT FT DIISUFID             255       276  
FT FT DIISUFID             286       336  
FT FT DIISUFID             295       319  
FT FT DISUFLD              311       332  
FT FT CARBOHYD             114       114  
FT FT CARBOHYD             233       233  
FT FT ACT_SITE              240       241  
  
ACT_SITE   296           297  
  
CONFLICT    142          142  
SEQUENCE    349 AA; 38851 MW; 1B7F97DCB0824E01 CRC64;  
                  G->A (IN REF. 2).
```

Query Match	7.18;	Score 214;	DB 1;	Length 349;
Best Local Similarity	40.68;	Pred. No. 7.1e-07;		
Matches	39;	Conservative	14;	Mismatches 41;
			Indels	2;
			Gaps	1;

QY 317 WHYPQRGCGMTFPARGCDGAARGFETYEACQACARGGDACVLPVAGQCRGWBRWA 376
::: ||| ||| : : ||| ||| ||| : ||
Db 246 YYNGASMACETFYQGGCLGNMNNFASKEKCLQTCRTIA--ACNLPYVQSGCRAFALIMA 303

Qy	377	YSPLLQCHPFYVGGCGEENGNHRESEDCAPV	412
Db	304	FDAAGKCIQFTYGGCKNGNMFYSECKEYCGVP	339
RESULT	5		
AMBP_HUMAN			
ID	AMBP_HUMAN	STANDARD:	PRF: 352 AA.
AC	P02760; P02759; P00977;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	13-AUG-1987 (Rel. 05, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN (PROTEIN HC) (COMPLEX-FORMING GLYCOPROTEIN HETEROGENEOUS IN CHARGE); INTER-ALPHA-TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].		
AM	OR ITIL OR HCP.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91214554; PubMed=1708673;		
RA	Veit H., Gebhard W.;		
RT	"Structure of the human alpha 1-microglobulin-bikunin gene.";		
RL	Biol. Chem. Hoppe-Seyler 371:1185-1196(1990).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=87040757; PubMed=2430261;		
RA	Kaumeier J.F., Polazzi J.O., Kotlick M.P.;		
RT	"The mRNA for a proteinase inhibitor related to the HI-30 domain of		
RT	inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin		
RL	(protein HC).";		
RN	Nucleic Acids Res. 14:7839-7850(1986).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=90336621; PubMed=1696200;		
RA	Dierra-Meirpour M., Bourguignon J., Sesshou R., Saller J.P.,		
RT	Leveillard T., Martin J.P.;		
RT	"Structural analysis of the human inter-alpha-trypsin inhibitor		
RL	light-chain gene.";		
RN	Eur. J. Biochem. 191:131-139(1990).		
RN	[4]		
RP	SEQUENCE OF 1-220 FROM N.A.		
RX	MEDLINE=86312901; PubMed=2428011;		
RA	Traboni C., Cortese R.;		
RT	"Sequence of a full length cDNA coding for human protein HC (alpha 1		
RL	microglobulin).";		
RN	Nucleic Acids Res. 14:6340-6340(1986).		
RN	[5]		
RP	SEQUENCE OF 20-202 (INDIVIDUAL WITH TUBULAR PROTEINURIA).		
RX	MEDLINE=84126849; PubMed=6198962;		
RA	Lopez C., Grubb A.O., Mendez E.;		
RT	"The complete amino acid sequence of human complex-forming		
RT	glycoprotein heterogeneous in charge (protein HC) from one		
RL	individual.";		
RN	Arch. Biochem. Biophys. 228:544-554(1984).		
RN	[6]		
RP	SEQUENCE OF 20-198 (VARIANT).		
RA	Lopez C., Grubb A.O., Mendez E.;		
RT	"human protein HC displays variability in its carboxyl-terminal amino		
RT	acid sequence.";		
RL	FEBS Lett. 144:349-353(1982).		
RN	[7]		
RP	SEQUENCE OF 20-198 (PATIENTS WITH TUBULAR PROTEINURIA).		
RX	MEDLINE=81184038; PubMed=6164372;		
RA	Takagi T., Takagi K., Kawai T.;		
RT	"Complete amino acid sequence of human alpha 1-microglobulin.";		
RL	Biochem. Biophys. Res. Commun. 98:997-1001(1981).		
RN	[8]		
RP	SEQUENCE OF 206-350.		

SEQUENCE OF 206-350.

RX MEDLINE=85225968; PubMed=2408638;
 RA Reisinger P., Hochstrasser K., Albrecht G.J., Lempart K., Sallier J.P.;
 RT "Human Inter-alpha-trypsin inhibitor: localization of the kunitz-type
 RT domains in the N-terminal part of the molecule and their release by a
 RT trypsin-like proteinase.";
 RL Biol. Chem. Hoppe-Seyler 366:479-483(1985).
 RN [9]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=82074265; PubMed=6171497;
 RA Hochstrasser K., Schoenberger O.U., Rossmannith I., Wachter E.;
 RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
 RT the inter-alpha-trypsin inhibitor. V. Attachments of carbohydrates in
 RT the human urinary trypsin inhibitor isolated by affinity
 RT chromatography.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 362:1357-1362(1981).
 RN [10]
 RP INHIBITORY SITE.
 RX MEDLINE=85225940; PubMed=3890890;
 RA Moril M., Travis J.;
 RT "The reactive site of human inter-alpha-trypsin inhibitor is in the
 RT amino-terminal half of the protein.";
 RL Biol. Chem. Hoppe-Seyler 366:19-21(1985).
 RN [11]
 RP STRUCTURE OF CARBOHYDRATES.
 RX MEDLINE=90306345; PubMed=1694784;
 RA Escribano J., Lopez-Otin C., Hjerpe A., Grubb A., Mendez E.;
 RT "Location and characterization of the three carbohydrate prosthetic
 RT groups of human protein HC.";
 RL FEBS Lett. 266:167-170(1990).
 RN [12]
 RP CHROMOPHORE.
 RX MEDLINE=91340714; PubMed=1714898;
 RA Escribano J., Grubb A., Calero M., Mendez E.;
 RT "The protein HC chromophore is linked to the cysteine residue at
 RT position 34 of the polypeptide chain by a reduction-resistant bond
 RT and causes the charge heterogeneity of protein HC.";
 RL J. Biol. Chem. 266:15758-15763(1991).
 RN [13]
 RP SEQUENCE OF 206-219, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
 RX MEDLINE=94229087; PubMed=7513643;
 RA Morelle W., Capon C., Balduyck M., Sautiere P., Kouach M.,
 RA Michalski C., Fournet B., Mizon J.;
 RT "Chondroitin sulphate covalently cross-links the three polypeptide
 RT chains of inter-alpha-trypsin inhibitor.";
 RL Eur. J. Biochem. 221:881-888(1994).
 RN [14]
 RP SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC2.
 RX MEDLINE=93232026; PubMed=7682553;
 RA Engblid J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
 RA Pizzo S.V., Hetta S.A.;
 RT "Presence of the protein-glycosaminoglycan-protein covalent cross-link
 RT in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
 RT 2/Bikunin.";
 RL J. Biol. Chem. 268:8711-8716(1993).
 RN [15]
 RP SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC3.
 RX MEDLINE=91093267; PubMed=1898736;
 RA Engblid J.J., Salvesen G., Hetta S.A., Thøgersen I.B.,
 RA Rutherford S., Pizzo S.V.;
 RT "Chondroitin 4-sulfate covalently cross-links the chains of the human
 RT blood protein pre-alpha-inhibitor.";
 RL J. Biol. Chem. 266:747-751(1991).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 230-339.
 RX MEDLINE=98227321; PubMed=9566199;
 RA Xu Y., Carr P.D., Guss J.M., Ollis D.L.;
 RT "The crystal structure of bikunin from the inter-alpha-inhibitor
 RT complex: a serine protease inhibitor with two kunitz domains.";
 RL J. Mol. Biol. 276:955-966(1998).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT

CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN.
 CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE. ADDITIONAL PROTEOLYTIC PROCESSING IN THE KIDNEY AND/OR
 CC URINE CAN PRODUCE FURTHER AMINO- AND CARBOXYL-END MODIFICATIONS
 CC IN ITS SEQUENCE.
 CC -1- SUBUNIT: INTER-ALPHA-TRYPsin INHIBITOR CONSIST OF A LIGHT CHAIN
 CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
 CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
 CC -1- PTM: ADDITION OF GLYCOSAMINOGLYCAN CHONDROITIN SULFATE, ALLOWS
 CC CROSS-LINKING BETWEEN THE DIFFERENT COMPONENTS.
 CC -1- MISCELLANEOUS: IN VITRO, THE FIRST TWELVE RESIDUES OF THE AMINO
 CC END OF THE INHIBITOR APPEAR TO HAVE A REACTIVE SITE CAPABLE OF
 CC INHIBITING THE ACTIVITY OF A NUMBER OF ENZYMES. ITS IN VIVO
 CC FUNCTION IS NOT KNOWN.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL; X54816; CAA38585.1; -;
 DR EMBL; X54817; CAA38585.1; JOINED.
 DR EMBL; X54818; CAA38585.1; JOINED.
 DR EMBL; X04225; CAA27803.1; -;
 DR EMBL; M88249; AAA59196.1; -;
 DR EMBL; M88165; AAA59196.1; JOINED.
 DR EMBL; M88243; AAA59196.1; JOINED.
 DR EMBL; M88244; AAA59196.1; JOINED.
 DR EMBL; M88246; AAA59196.1; JOINED.
 DR EMBL; M88247; AAA59196.1; JOINED.
 DR EMBL; X04494; CAA28182.1; -;
 DR EMBL; X54817; CAA38586.1; -;
 DR PIR; A03217; HCHV.
 DR PIR; A25303; A25303.
 DR PIR; S13433; S13433.
 DR PIR; S10717; S10717.
 DR PDB; 1B1K; 16-MAR-99.
 DR SWISS-2DPAGE; P02760; HUMAN.
 DR MIM; 176870; -;
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR002345; Lipocalin.
 DR InterPro; IPR000566; Lipocalin_cytrBP.
 DR Pfam; PF00014; Kunitz_BPTI_2.
 DR Pfam; PF00061; Lipocalin_1.
 DR PRINTS; PR00179; Lipocalin.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS0279; BPTI_KUNITZ_2; 2.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;

Query Match 7.18; Score 213; DB 1; Length 352;
 Best Local Similarity 40.2%; Pred. NO. 8.3e-07;
 Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

GY CGTSPPLVLMHYDPOGGGCTFPARGCDAARGEFFTEACQACARGPDACVLPVQ 365
 DB 240 CMGMSRYF---YNGTMACTEFGYCGMGNGNFFTEKELQCRIVA--ACNLPIVR 293
 GY 366 GPCRGPWRMAYSPPLQOCHPFFVYGCCEGNGNFFHSRSCEDACVP 412
 DB 294 GPCRFTDMAFADVAKGCVLFPGYCGCGNGKFKFSKRECYGCV 340


```

Db 166 REVALGVCIPDPAITFMDRCGVCGEQDPVPTPLSRARAVLQEEBSGAGCPVTWFS 225
Oy 305 -----TGPISPLVLMHNDPQRCGCMTPARCCDGAARGFEYEAQACARGP 355
Db 226 KKAASCOLDVSOGDPLGFKRYFYNGTSMACETFLYGCMNGNNFLSEKEDLOTCTA--T 283
Oy 356 GDACVLPAVPGRCRGMEDRMVAVSPLOQCHPFVYGGCGGNGNNFHSRSCEDACVP 412
Db 284 VEAQNLPIVQPCRSYIOLMAFDAYKGCVRFSIGCGKNGKRYSEKCEKCYGIP 340

RESULT 7
IATR_HORSE STANDARD: PRT: 123 AA.
AC P04365;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTER-ALPHA-TRYPsin INHIBITOR (ITI) (HI-14) (INHIBITORY FRAGMENT OF ITI) (FRAGMENT).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RX MEDLINE=85225967; PubMed=2408637;
RA Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-alpha-trypsin inhibitor. X. The amino-acid sequences of the trypsin-released inhibitors from horse and pig inter-alpha-trypsin inhibitors."
RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).
CC -1- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER LIMITED PROTEOLYSIS WITH TRYPsin, CONTAINS TWO HOMOLOGOUS DOMAINS. WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPsin, THE FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND NOT AT ALL WITH PANCREATIC ELASTASE.
CC -1- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I. INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
DR PIR: A01210; TIMOBI.
DR HSSP; P10646; IADZ.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR SMART: SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT DOMAIN 1 1 55 BPTI/KUNITZ INHIBITOR 1.
FT NON_TER 1 1 BPTI/KUNITZ INHIBITOR 2.
FT DOMAIN 61 111
FT DISULFID 5 55
FT DISULFID 14 38
FT DISULFID 30 51
FT DISULFID 61 111
FT DISULFID 70 94
FT DISULFID 86 107
FT ACT_SITE 15 16
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN, ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPsin).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
FT NON_TER 123 123
SEQUENCE 123 AA; 13510 MW; CE1A9120774411D5 CRC64;

Query Match 6.98; Score 208; DB 1; Length 123;
Best Local Similarity 38.18; Pred. No. 6,4e-07;
Matches 40; Conservative 14; Mismatches 49; Indels 2; Gaps 1;

```

```

Oy 308 GPTSPHLVLMHYDPQRCGCMTPARCCDGAARGFEYEAQACARGDACLPAVQCP 367
Db 12 GPCGLMISRYFLNGTSMACETFLYGCLANGNNFASQKEDLOTCTA--ACNLPIVQCP 69
Oy 368 CRGWEPRMAYSPDLLQCHPFVYGGCGGNGNNFHSRSCEDACVP 412
Db 70 CNAFIRLMAFDAAQKCVLFTYGGCRGNGNKRYSQKECKEYCGIP 114

RESULT 8
IATR_SHEEP STANDARD: PRT: 123 AA.
AC P13371;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTER-ALPHA-TRYPsin INHIBITOR (ITI) (GIK-14) (INHIBITORY FRAGMENT OF ITI) (FRAGMENT).
OS Ovis aries (Sheep), and Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940, 9925;
RN [1]
RP SEQUENCE.
RC SPECIES=Sheep;
RX MEDLINE=87299012; PubMed=2441725;
RA Rasp G., Hochstrasser K., Wachter E., Reisinger P.W.M.;
RT "The amino-acid sequence of the trypsin-released inhibitor from sheep inter-alpha-trypsin inhibitor."
RL Biol. Chem. Hoppe-Seyler 368:727-731(1987).
RN [2]
RP SEQUENCE.
RC SPECIES=C.hircus;
RX MEDLINE=90105540; PubMed=2481505;
RA Rasp G., Hochstrasser K., Gerl C., Wachter E.;
RT "Primary structure of a proteinase inhibitor released from goat serum inter-alpha-trypsin inhibitor."
RL Biochim. Biophys. Acta 999:335-337(1989).
CC -1- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER LIMITED PROTEOLYSIS WITH TRYPsin, CONTAINS TWO HOMOLOGOUS DOMAINS. WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPsin, THE FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND NOT AT ALL WITH PANCREATIC ELASTASE.
CC -1- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I. INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
DR PIR: A29652; A29652.
DR HSSP; P10646; IADZ.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR SMART: SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT DOMAIN 1 1 55 BPTI/KUNITZ INHIBITOR 1.
FT NON_TER 1 1 BPTI/KUNITZ INHIBITOR 2.
FT DOMAIN 61 111
FT DISULFID 5 55
FT DISULFID 14 38
FT DISULFID 30 51
FT DISULFID 61 111
FT DISULFID 70 94
FT DISULFID 86 107
FT ACT_SITE 15 16
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN, ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPsin).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
FT NON_TER 123 123
SEQUENCE 123 AA; 13686 MW; 295038173F22D2D1 CRC64;

```

Query Match 6.8%; Score 205; DB 1; Length 123;
 Best Local Similarity 37.1%; Pred. No. 9.9e-07;
 Matches 39; Conservative 15; Mismatches 49; Indels 2; Gaps 1;

QY 308 GPTSPHLVLMHYDPORGCMTPFANGCGAARGFETACCOACARGDACPVP 367
 DB 12 GPCLCMKRYFTNGTSMACEFYGGCMNGNMFSEKELQTCR--TYQACNLPYRGP 69
 QY 368 CGWPEPMAYSPDLQOCHPEFYGGCEGNGNMFHSECEADCPVP 412
 DB 70 CRAIGELMAFDVAVKGCVRFLYGGCNGNGNFIQKCKEYCGIP 114

RESULT 9
 ID TPEL_RABIT STANDARD; PRT; 300 AA.

AC P19761; Q28828;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)
 GN TFPI.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_Taxid=9986;
 RN 1
 RP SEQUENCE FROM N.A.
 RC TISSUE-Liver;
 RA MEDLINE=91057146; PubMed=2136251;
 RT Wesselschmidt R.L., Girard T.J., Broze G.J. Jr.;
 RL "CDNA sequence of rabbit lipoprotein-associated coagulation
 RL Nucleic Acids Res. 18:6440-6440(1990).
 RN [2]
 RP REVISIONS TO 72; 211 AND 218.
 RC TISSUE-Liver;
 RX MEDLINE=92335027; PubMed=1630940;
 RA Warn-Cramer B.J., Broze G.J. Jr., Komives E.A.;
 RT "CDNA sequence of rabbit tissue factor pathway inhibitor";
 RL Nucleic Acids Res. 20:3548-3548(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Lung;
 RX MEDLINE=93276427; PubMed=8503123;
 RA Belacqua A., Kuppuswamy M.N., Birktoft J.J., Bajaj S.P.;
 RT "Revised CDNA sequence of rabbit tissue factor pathway inhibitor";
 RL Thromb. Res. 69:547-553(1993).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC A OUTERNARY X(A)/LACT/VI(A)/TF COMPLEX, IT POSSESSES AN
 CC ANTIHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 CC LIPOPROTEINS IN PLASMA.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: X54708; CA38515.1; ALT_SEQ.
 CC EMBL: S61902; AA26836.1; -.
 CC PIR: S12143; S12143.
 CC HSSP: P10646; ITPX.
 CC InterPro: IPR002223; Kunitz_BPTI.

DR Pfam: PF00014; Kunitz_BPTI, 3.
 DR PRINTS: PR00759; BASICPASE.
 DR SMART: SM00131; KU; 3.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 3.
 KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 KW Signal.

FT CHAIN 1 24
 FT CHAIN 25 300
 FT DOMAIN 50 100
 FT DOMAIN 121 171
 FT DOMAIN 213 263

FT DISULFID 50 100
 FT DISULFID 59 83
 FT DISULFID 75 96
 FT ACT_SITE 60 61
 FT DISULFID 121 171
 FT DISULFID 130 154
 FT DISULFID 146 167
 FT ACT_SITE 131 132
 FT DISULFID 213 263
 FT DISULFID 222 246
 FT DISULFID 238 259
 FT ACT_SITE 223 224
 FT CARBOHYD 141 141
 FT CARBOHYD 191 191
 FT CARBOHYD 252 252
 FT CONFLICT 31 31
 FT CONFLICT 269 272
 SQ SEQUENCE 300 AA; 34435 MW; A08DE3537708CA6 CRC64;

Query Match 6.8%; Score 204; DB 1; Length 300;
 Best Local Similarity 31.1%; Pred. No. 2.6e-06;
 Matches 50; Conservative 20; Mismatches 71; Indels 20; Gaps 3;

QY 270 GLLRADPEPLSYVQREPAR-----DAAPSPAPACLDPDVQACTGPTSPHLVLMHYDPORG 324
 DB 16 GLVPA--FVSSAAEDDEFTNTIDIKPPLQKPTSPFCAMKVDQPCRAVYIKRFFENLTTH 73
 QY 325 GCMTPFANGCGAARGFETACCOACAR-----PGDACPVP 371
 DB 74 QCEEFYGGCEGNGNMFSELECKEKCARDYPKMTTKLTFQKRPDCELEDDPGICRGY 133
 QY 372 EPRMAYSPDLQOCHPEFYGGCEGNGNMFHSECEADCPVP 412
 DB 134 ITRFTYNNQSKQRCERFYGGCLGNLNFSELECKKTCENP 174

RESULT 10
 ID AMBP_MOUSE STANDARD; PRT; 349 AA.

AC 007456; O61294;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMBP PROTEIN PRECURSOR (CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)).
 GN AMBP OR ITIL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-Liver;
 RX MEDLINE=93363639; PubMed=7689339;
 RA Chan P., Sailer J.P.;
 RT "Mouse alpha-1-microglobulin/dikunin precursor: CDNA analysis, gene
 evolution and physical assignment of the gene next to the orosomucoid

RX MEDLINE=90036996; PubMed=2553722;
 RA Novotny W.F., Girard T.J., Miletich J.P., Broze G.J. Jr.;
 RT "Purification and characterization of the lipoprotein-associated
 RT coagulation inhibitor from human plasma.";
 RT J. Biol. Chem. 264:18832-18837(1989).
 RN [6]
 RP INHIBITORY SITES.
 RX MEDLINE=89181950; PubMed=2927510;
 RA Girard T.J., Warren L.A., Novotny W.F., Likert K.M., Brown S.G.,
 RA Miletich J.P., Broze G.J. Jr.;
 RT "Functional significance of the Kunitz-type inhibitory domains of
 RT lipoprotein-associated coagulation inhibitor.";
 RT Nature 338:518-520(1989).
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=96224851; PubMed=8639592;
 RA Nakahara Y., Miyata T., Hamuro T., Funatsu A., Miyagi M.,
 RA Tsunasawa S., Kato H.;
 RT "Amino acid sequence and carbohydrate structure of a recombinant
 RT human tissue factor pathway inhibitor expressed in Chinese hamster
 RT ovary cells: one N- and two O-linked carbohydrate chains are located
 RT between Kunitz domains 2 and 3 and one N-linked carbohydrate chain is
 RT in Kunitz domain 2.";
 RT Biochemistry 35:6450-6459(1996).
 RN [8]
 RP REVIEW.
 RX MEDLINE=91104709; PubMed=2271516;
 RA Broze G.J. Jr., Girard T.J., Novotny W.F.;
 RT "Regulation of coagulation by a multivalent Kunitz-type inhibitor.";
 RT Biochemistry 29:7539-7546(1990).
 RN [9]
 RP STRUCTURE BY NMR OF 121-182.
 RX MEDLINE=97342711; PubMed=9199408;
 RA Burgering M.J., Orhons L.P., van der Doelen A., Mulders J.,
 RA Theunissen H.J., Grootenhuys P.D., Bode W., Huber R., Stubbs M.T.;
 RT "The second Kunitz domain of human tissue factor pathway inhibitor:
 RT cloning, structure determination and interaction with factor Xa.";
 RT J. Mol. Biol. 269:395-407(1997).
 RN [10]
 RP FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC A QUATERNARY X(A)/LACI/VIII(A)/TF COMPLEX. IT POSSESSES AN
 CC ANTITHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 CC LIPOPROTEINS IN PLASMA.
 CC [11] TISSUE LOCATION: SECRETED.
 CC [12] TISSUE SPECIFICITY: MOSTLY IN ENDOTHELIAL CELLS.
 CC [13] DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC [14] PTM: O-GLYCOSYLATED.
 CC [15] SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 DR EMBL: J03225; AAA52022.1; -;
 DR EMBL: M58650; AAA59480.1; -;
 DR EMBL: M58644; AAA59480.1; JOINED.
 DR EMBL: M58645; AAA59480.1; JOINED.
 DR EMBL: M58646; AAA59480.1; JOINED.
 DR EMBL: M58647; AAA59480.1; JOINED.
 DR EMBL: M58648; AAA59480.1; JOINED.
 DR EMBL: M58649; AAA59480.1; JOINED.
 DR EMBL: M59489; AAA59526.1; -;
 DR EMBL: M59493; AAA59526.1; JOINED.
 DR EMBL: M59494; AAA59526.1; JOINED.
 DR EMBL: M59495; AAA59526.1; JOINED.
 DR EMBL: M59496; AAA59526.1; JOINED.
 DR EMBL: M59497; AAA59526.1; JOINED.
 DR EMBL: M59498; AAA59526.1; JOINED.
 DR PIR: A28650; TIRHGK.

DR PIR: A43315; A34315.
 DR PIR: A60433; A60433.
 DR PIR: S03903; S03903.
 DR PDB: 1ADZ; 25-FEB-98.
 DR PDB: 1TFX; 21-JAN-98.
 DR GlycoSuiteDB; P10646; -.
 DR MIM; 152310; -.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR Pfam; PF00014; Kunitz_BPTI_3.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00131; KU; 3.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE; PS0279; BPTI_KUNITZ_2; 3.
 KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 FT SIGNAL 1 28
 FT CHAIN 29 304
 FT DOMAIN 54 104
 FT
 FT DOMAIN 125 175
 FT
 FT DOMAIN 217 267
 FT DISULFID 54 104
 FT DISULFID 63 87
 FT DISULFID 79 100
 FT ACT_SITE 64 65
 FT DISULFID 125 175
 FT DISULFID 134 158
 FT DISULFID 150 171
 FT ACT_SITE 135 136
 FT DISULFID 217 267
 FT DISULFID 226 250
 FT DISULFID 242 263
 FT ACT_SITE 227 228
 FT CARBOHYD 145 145
 FT CARBOHYD 202 202
 FT CARBOHYD 203 203
 FT CARBOHYD 195 195
 FT MUTAGEN 64 64
 FT MUTAGEN 135 135
 FT MUTAGEN 227 227
 SO SEQUENCE 304 AA; 35015 MM; 5281832875844FE CRC64;
 Query Match 6.4%; Score 194.5; DB 1; Length 304;
 Best Local Similarity 27.3%; Pred. NO. 1.1e-05;
 Matches 50; Conservative 20; Mismatches 58; Indels 55; Gaps 5;
 QY 263 CTARNAAGLRADPPLSVQREPARDAAPSLPAPAECL--PDVQACTGPTSPHLVLMHYD 320
 DB 104 CTRDANARIRK-----TTLQOE-----KDFCFLEDEPDGICG---YITRTFYN 144
 QY 321 PORGCGMTFPARGGCGDGAARGFETYEACQACARPG----- 356
 DB 145 NOTQCECFKRGKGLGMNMFETILECKNICEDGPNFGQVDNYGTOLNAYNNSLTPQSTR 204
 QY 357 -----DACVLPAVQGPCRCGMEPRMAYSPLOQCCHFYVGGGNGNMFHSRESQ 406
 DB 205 VPSLFEFHGSPSWLTPADRLCLCRANENREYFNVYICKRCRFKSGCGGNENFTSKQEC 264
 QY 407 DAC 409
 DB 265 RAC 267
 RESULT 12
 SPT2_HUMAN
 ID SEPT2_HUMAN STANDARD; PRT; 252 AA.
 AC 043291; 000271; 014895;
 DT 20-AUG-2001 (rel. 40; Created)
 DT 20-AUG-2001 (rel. 40; Last sequence update)
 DT 20-AUG-2001 (rel. 40; Last annotation update)
 DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR

DE ACTIVATOR INHIBITOR TYPE 2 (HA1-2) (PLACENTAL BIKUNIN).
 CN SPRINT2 OR HA12 OR KOP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96010584; PubMed=9346890;
 RA Kawaguchi T., Qin L., Shimomura T., Kondo J., Matsumoto K., Denda K.,
 RA Kitamura N.;
 RT "Purification and cloning of hepatocyte growth factor activator
 RT inhibitor type 2, a Kunitz-type serine protease inhibitor.";
 RL J. Biol. Chem. 272:27558-27564(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-74.
 RX MEDLINE=97277372; PubMed=9115294;
 RA Marlor C.W., Delaria K.A., Davis G., Muller D.K., Greve J.M.,
 RA Tamburini P.P.;
 RT "Identification and cloning of human placental bikunin, a novel serine
 RT protease inhibitor containing two Kunitz domains";
 RL J. Biol. Chem. 272:12202-12208(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Pancreatic cancer;
 RX MEDLINE=98094245; PubMed=9434156;
 RA Mueller-Pillasch F., Wallrapp C., Bartels K., Varga G., Friess H.,
 RA Buechler M., Adler G., Gress T.M.;
 RT "Cloning of a new Kunitz-type protease inhibitor with a putative
 RT transmembrane domain overexpressed in pancreatic cancer.";
 RL Blochim. Biophys. Acta 1395:88-95(1998).
 CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR. ALSO INHIBITS PLASMIN,
 CC PLASMA AND TISSUE KALLIKREIN, AND FACTOR XIA.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, KIDNEY, PANCREAS,
 CC PROSTATE, TESTIS, THYMUS, AND TRACHEA.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB006534; BAA25024.1; -
 DR EMBL: U78095; AAC02781.1; -
 DR EMBL: AF037205; AAB84031.1; -
 DR HSP: P05067; ITAM.
 DR MIM: 605124; -
 DR InterPro: IPR002223; Kunitz-BPTI.
 DR Pfam: PF00014; Kunitz-BPTI; 2.
 DR PRINTS: PRO0759; BASICPTASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 KW Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;
 KW Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.
 FT DOMAIN 28 197 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 198 218 POTENTIAL.
 FT DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
 FT DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
 FT DISULFID 38 88 BY SIMILARITY.
 FT DISULFID 47 71 BY SIMILARITY.
 FT DISULFID 63 84 BY SIMILARITY.
 FT ACT_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 133 183 BY SIMILARITY.

FT DISULFID 142 166 BY SIMILARITY.
 FT DISULFID 158 179 BY SIMILARITY.
 FT ACT_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 3 3 O -> H (IN REF. 3).
 FT CONFLICT 11 11 R -> P (IN REF. 1).
 FT CONFLICT 53 53 R -> K (IN REF. 3).
 FT CONFLICT 240 240 D -> H (IN REF. 3).
 SQ SEQUENCE 252 AA; 28228 MW; A7D3360CDECA82B CRC64;
 Query Match 6.4%; Score 192; DB 1; Length 252;
 Best Local Similarity 28.5%; Pred. No. 1.3e-05;
 Matches 45; Conservative 13; Mismatches 54; Indels 46; Gaps 4;
 QY 299 CLPDVQACTGPTSHLVNHPDQRCGCTFPARGCGAARGETFEACQACA----- 352
 DB 38 CL-VSKVYGRGRASRPWRMYNTDSSCOLFYVGGCDGNSNNYLTKECLLKCATYENA 95
 QY 353 -----KPPGDA-----CULPVGQPCRCWEPRAWYSPLL 381
 DB 96 TGDLATSRNAADSSVPSAPRRQDSEDDHSSDMFNYEEYCTANAVTGPCRFASFPWYDVER 155
 QY 382 QOCHPVGCGEGNGNHNFSRSECDAC-----PVP 412
 DB 156 NSCNNTFYGGCRKNKNSYSEECMLRCRQDENPLP 193
 RESULT 13
 SEPT2_MOUSE STANDARD; PRT; 252 AA.
 AC Q9W003; Q9W004; Q9W005;
 DT 20-AUG-2001 (Rel. 40; Created)
 DT 20-AUG-2001 (Rel. 40; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR
 DE ACTIVATOR INHIBITOR TYPE 2) (HA1-2).
 GN SPRINT2 OR HA12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=BAH/C;
 RX MEDLINE=99160423; PubMed=10049781;
 RA Itoh H., Kataoka H., Hamasuna R., Kitamura N., Koono M.;
 RT "Hepatocyte growth factor activator inhibitor type 2 lacking the first
 RT Kunitz-type serine proteinase inhibitor domain is a predominant
 RT product in mouse but not in human.";
 RL Biochem. Biophys. Res. Commun. 255:740-748(1999).
 CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: ISOFORM 2 IS MORE PREDOMINANTLY EXPRESSED THAN
 CC ISOFORM 1.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF099016; AAD22172.1; -
 DR EMBL: AF099019; AAD22173.1; -
 DR EMBL: AF099020; AAD22174.1; -
 DR HSP: P05067; ITAM.

DR MGD; MG1:1338031; Spint2.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS02279; BPTI_KUNITZ_2; 2.
 KW Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;
 KW Signal; Alternative splicing.
 FT SIGNAL 1 27
 FT CHAIN 28 252
 FT DOMAIN 28 197
 FT TRANSMEM 198 218
 FT DOMAIN 219 252
 FT DOMAIN 133 88
 FT DOMAIN 183 88
 FT DISULFID 38 88
 FT DISULFID 47 71
 FT DISULFID 63 84
 FT ACT_SITE 48 49
 FT DISULFID 133 183
 FT DISULFID 142 166
 FT DISULFID 158 179
 FT ACT_SITE 143 144
 FT CARBOHYD 57 57
 FT CARBOHYD 94 94
 FT VARSPPLIC 37 93
 FT VARSPPLIC 114 128
 FT VARSPPLIC 129 252
 FT VARSPPLIC 252 AA; 27914 MW; B2FFB8692404F8 CRC64;
 SQ SEQUENCE

Query Match
 Best Local Similarity 6.48; Score 192; DB 1; Length 252;
 Matches 54; Conservative 20; Mismatches 87; Indels 46; Gaps 5;

QY 263 CTARNAAGLR--ADPLSYVQREPARAASIPAPACLDYVQACGPGPSPLVLWHYD 320
 DB 5 CELRGRLALLVASLILS-----GAQVASRELVHESC--GVSRYVKGCRAISIPRWYN 57
 QY 321 PORGCGMTPFARGCDGAARGFETYEAQOACA-----RCGQDA----- 358
 DB 58 ITDSCQPFYVGGCEGNGNNVSKKECLDKCAGTENTTDMANRRGAGSSVLSVPRKQ 117
 QY 359 -----CYLPANQSGRGMERPMATSPILQOCHPVGCGGNGNHSRE 403
 DB 118 SAEDLSAEINVEYECVPKAVTGPCRAFPWYDTEKNSCISFTYIGCGRGNKNSYLSQE 177
 QY 404 SCEDACFPVPTPCRCRCLRSKLALSL 430
 DB 178 ACMQHCQSKQMHPFLTPGLKAVILVGL 204

RESULT 14
 YN81_CAEEL STANDARD; PRT; 1416 AA.
 ID YN81_CAEEL
 AC 003610;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOHETICAL 316.1 KDA PROTEIN ZC84.1 IN CHROMOSOME III.
 GN ZC84.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favella A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Weinstock L., Wilkinson-Sprat J.,
 RA Wohldman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RA elegans.";
 RA Nature 368:32-38(1994).
 [2]
 REVISIONS.
 RC STRAIN-BRISTOL N2;
 RA Jones S.J.M.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: STRONG, TO D1044.3.
 CC -1- SIMILARITY: CONTAINS 5 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z19157; CAA79569.1; -
 DR PIR; S28291; S28291.
 DR HSSP; P07204; 2ADX.
 DR Wormpep; ZC84.1; CE15020.
 DR InterPro; IPR002899; EB.
 DR InterPro; IPR000561; EGF-1like.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR Pfam; PF01683; EB; 3.
 DR Pfam; PF00014; Kunitz_BPTI; 5.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART; SM00131; KU; 5.
 DR SMART; SM00289; WRI; 13.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE; PS02279; BPTI_KUNITZ_2; 5.
 DR PROSITE; PS01186; EGF_2; UNKNOWN.1.
 KW Hypothetical protein; Serine protease inhibitor; Repeat.
 FT DOMAIN 212 266
 FT DOMAIN 337 387
 FT DOMAIN 434 484
 FT DOMAIN 538 590
 FT DOMAIN 646 698
 FT SEQUENCE 1416 AA; 152986 MW; 531CACEL1CR22F70D CRC64;

Query Match
 Best Local Similarity 6.24; Score 188; DB 1; Length 1416;
 Matches 104; Conservative 44; Mismatches 181; Indels 234; Gaps 22;

QY 9 PLLILLRLTSGAGL-----LPGLSGHPGVCNPOLSPNLMWVDAQSGCERCSDQOCAA 61
 DB 112 PLFYFCALPEGAGFRISLNDSPSMGAFVCGT-----VGSATIEPV-----QGLHG 159
 QY 62 AEKC-----CIVACGLHSCVAAFPGSPAPPTTAASCGFVCPQOQSDC-- 105
 DB 160 GGACWMPSDGEYIGDYVDICHT-SPHINLAEKFPQSNVAPY---ADGVCCPSALACIQ 214
 QY 106 -----DINDGGPV-----CR--CRDRCK-- 122
 DB 215 PQLGPNSEPRRWYNSVTGCOQPFMDSCGATKAYSPNNKFTIQHCSEYCRDTCORG 274
 QY 123 -----EPSTCASDGLIYNYRCYMDAEACIRGLHLIIVPCKHLISWPPSPGP 170
 DB 275 PQYTESAVHSERPTSCAA-----SVSCANDPQCAIASHQHLCC----- 314
 QY 171 PETTARPLPGAAPVPALYSSPSQAVQVGTASLHCDVSGRPPAVTWKESHORENLI 230

```

Db 315 -----PRFGSV-----CSKAGGRPYDI-----QPRSTI 337
Oy 231 MRPDQMGVNVVTSIGOLVY---NARPD-----AGLYT-----CTARNAGILIR 273
Db 338 HHP-----GPMVVTGKRESIFVYDPTGRCODEFVYKAGAGNNYNNLSKHCEMYCARLOCE 393
Oy 274 ADFPLSVQREPARDAAPSIAPACLPDVOAC-----TGTPSPHLYVMHND 320
Db 334 RGSPLRIGEEARCONNACPSHCEKADGVCCPRKOTICAPLRIGCTENVRKYWN 453
Oy 321 PORGCMTFPGARCDGAARGFETEACQ-----ACARCP 355
Db 454 ATRROCMFEYTCGCGNDNPNFSIMDCNFKNAIPKCIQOAVKDMFGNFVTCNSGM 513
Oy 356 G----DACVLPVAVGPC-----RGMEPRMAYSPLLQOCHFPVYGGC 392
Db 514 GCPAVNECYFDSQWCCPTKAFCTSLMTDSCIQCGAGSTFKYYNPTQNCESFOYNGC 573
Oy 393 EGNGNNFHSRESCDACPVRTPP 415
Db 574 DGNMNFANRDACESTCYGCP 596

RESULT 15
TFPI_MOUSE STANDARD: PRT: 306 AA.
AC 054819;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)
DE (EPI).
GN TFPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=98152575; PubMed=9493581;
RA Chang J.-Y., Monroe D.M., Oliver J.A., Liles D.K., Roberts H.R.;
RT "Cloning, expression, and characterization of mouse tissue factor
RT Pathway Inhibitor (TFPI).";
RL Thromb. Haemost. 79:306-309(1998).
CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
CC MAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
CC A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN
CC ANTIHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
CC LIPOPROTEINS IN PLASMA (BY SIMILARITY).
CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF004833; AAC40035.1; -
DR HSSP; P10646; ITEX.
DR MGD; MGI:1095418; TFPI.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI; 3.
DR SMART; SM00131; KU; 3.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 3.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 3.

```

```

KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
KM Signal.
FT SIGNAL. 1 28 BY SIMILARITY.
FT CHAIN 29 306 TISSUE FACTOR PATHWAY INHIBITOR.
FT DOMAIN 50 100 BPTI/KUNITZ INHIBITOR 1
FT (VII(A)/TISSUE FACTOR BINDING SITE).
FT DOMAIN 121 171 (FACTOR X(A) INHIBITOR 2
FT (BPTI/KUNITZ INHIBITOR 3.
FT BPTI/KUNITZ INHIBITOR 3.
FT DOMAIN 225 275 BY SIMILARITY.
FT DISULFID 50 100 BY SIMILARITY.
FT DISULFID 59 83 BY SIMILARITY.
FT DISULFID 75 96 BY SIMILARITY.
FT ACT SITE 60 61 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 121 171 BY SIMILARITY.
FT DISULFID 130 154 BY SIMILARITY.
FT DISULFID 146 167 REACTIVE BOND (BY SIMILARITY).
FT ACT SITE 131 132 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 225 275 BY SIMILARITY.
FT DISULFID 234 258 BY SIMILARITY.
FT DISULFID 250 271 BY SIMILARITY.
FT ACT SITE 235 236 REACTIVE BOND (BY SIMILARITY).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 306 AA; 34987 MW; D3EA3297EAB6A359 CRC64;

```

Query Match 6.2%; Score 186.5; DB 1; Length 306;
 Best Local Similarity 31.8%; Pred. No. 3.4e-05;
 Matches 41; Conservative 17; Mismatches 54; Indels 17; Gaps 2;

```

Oy 303 VVACTGPTSPHLYVMHYDQRGCMTFPARGCDGAARGFETEACQACARPG----- 356
Db 52 MKADGPGCKAMIRSYFFNMVTHQCEFIYGGEGENRPRDTLECKKTCI--PGYEKTAV 109
Oy 357 -----DACVLPVAVGPCRGMEPRMAYSPLLQOCHFPVYGGCGNGNNFHSRESCD 407
Db 110 KAASGAERDPFCFLBEDPDLGCRGYMKRYLYNNQTCERFVYGGCLGNMNNFETIDECK 169
Oy 408 ACPVPRTPP 416
Db 170 ICENPVHSP 178

```

Search completed: February 26, 2002, 01:21:20
 Job time: 660 sec


```

QY 212 RPPAVTWK-----QSHORENLIIMRPPDMGNVYVTSIGQVLYNAREDA----- 258
Db 1327 AQPAAHQEQAOAPAOAPQAPQPSNIVSPPOQASAPVVP-----NSKORDACHLVN 1379
QY 259 -----GLYTCARNAAGLIRADPLSYVOREPARDAPSTPA 295
Db 1380 DQRCAGAFMSWYEVATGSCVTFKTYGCGNANRFASKDOQESLCVYPASBASAGIDG 1439
QY 296 PA--ECLIPVOACTGPTSPHLYLMHDDPQGGCMTPFARGCDGAARGFETEACQOACAR 353
Db 1440 AAGINSVCDEAKDGTCTNFTVTKWYNKADGICNRFHGGCGQTNNRDNDOCKAAC-Q 1498
QY 354 GPDGACVLPAYOGPCRGMPFMAVSPLLQCGHPFYVYGCSENGNHNFSRSCEDACP 410
Db 1499 NHRKDAQQLKVOGPGCSGHSYTYNTASHOCETFEYGGCLGNTNRFATIECCARCP 1555

```

RESULT 2

```

726063
Hypothetical protein W01F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26063
R:Cummings, P.
Submitted to the EMBL Data Library, March 1997
A:Reference number: Z20145
A:Accession: T26063
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2225 <MIL>
A:Cross-references: EMBL:Z92815; PIDN:CAB07294.1; GSPDB:GN00023; CESP:W01F3.3
A:Experimental source: clone W01F3
A:Gene: CESP:W01F3.3
A:Map position: 5
A:Introns: 33/1; 56/1; 100/1; 142/3; 271/3; 451/1; 525/3; 774/1; 1093/1; 1178/1; 1221/1;

```

```

Query Match
Best Local Similarity 7.8%; Score 236.5; DB 2; Length 2225;
Matches 123; Conservative 46; Mismatches 218; Indels 229; Gaps 28;

```

```

QY 28 SHPGVCPNQLSPNLMVDAQSTCEP-----ECSRDDCAAAKCCGCTINVC-----GLHSCVAA 78
Db 467 SLPGCHGHSFQRYFNEDSQKCEQPTYSGGCGNMYESRACEDRCAPPVGLPKCEI- 525
QY 79 RPPGSP-----AAPTTAASCEGFVCPQGGSDCDIMDGPVPC-----R 115
Db 526 ---GEPKTKIGVPVNCATKD--CP--SGYRCVYQHSVCCPENKNVYGLQTSGARATR 578
QY 116 CRDCKEKEPSTCASDGLITY--NRC-----YMDAE--ACLGLH- 151
Db 579 CSLPREKGP--CDKYLRFYFNADLNECKYFFWGCCEGNQNNFERVEDECESAC--GYCK 633
QY 152 -----LHIVPCKHVLISM-----PPSSP--GPPETTRAPTPGAP 183
Db 634 SGVNNRNTETRTTQGRITPNCGLKSMETEDDEHNAVPTTTPAPSPVYVSTQBAR 693
QY 184 VPAPVYSSPPOAVOYGTATLHCDVSGRPRAVYWEKOSHORENLIIMRPPDMGNVYV 243
Db 694 VP---TTPRPAVAVQTTTTRASRLTETTRVAVKTYVEDEEEEEEV- 736
QY 244 SIGQVLYNAREDAAGLYCTARNAAGLIRADPLSYVOREPARDAPSTPA 300
Db 737 -----EEOEDG-----KEEPPPLHVPDPVSOQNTVILGG 765
QY 301 -----PDVOACTGPTSP-----HLVLMHYDPQGGCTTPFARGCDGAARGFETEACQ 349
Db 766 IDTTTDSVNRCLHPRDGCNGCGVRFWFDEKKNCDVFYTTGCGQNGNNSKECMA 825
QY 350 ACARGP-----GDAACVLPAYOGPCRGMPFMAVSPLLQCGHPFYVYGCSENGNHN 400
Db 826 ICHKPEPPSATPDPFSQVCSNDVDAGECNGYERFAFDALADQCRAPTYGGCGGNGNNA 885

```

```

QY 401 SRSCEDACPVP--RTPPCRAC-----RLRSKALALCNS-----D 434
Db 886 TMOECSRVCVAMAKKSPATATCEADLEVEGCAGVTSRFAFDKSIACRSFTYGGCGNAN 945
QY 435 FAIVGRLTE-----VLEPEPAAGIARVALEDV-----LKDDKMGKLGCTKYLEVTL 484
Db 946 FATHLOCTNRCVNRGVCPEPPAC-----DINROQLVNDRSGCF----- 984
QY 485 MDWACPENMTAGDGP 500
Db 985 ---CSCPVPVQASPP 996

```

RESULT 3

```

719734
Hypothetical protein C34F6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T19734
R:White, S.
Submitted to the EMBL Data Library, November 1996
A:Reference number: Z19171
A:Accession: T19734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1043 <MIL>
A:Cross-references: EMBL:Z81479; PIDN:CAB03944.1; GSPDB:GN00028; CESP:C34F6.1
A:Experimental source: clone C34F6
A:Gene: CESP:C34F6.1
A:Map position: X
A:Introns: 17/3; 62/1; 129/3; 181/1; 292/1; 348/1; 404/1; 665/1; 769/1; 876/1; 977/3

```

```

Query Match
Best Local Similarity 7.2%; Score 218; DB 2; Length 1043;
Matches 100; Conservative 46; Mismatches 155; Indels 178; Gaps 20;

```

```

QY 53 CSRRDDCAAAKCCINVCGLHSCVAFRPPSPAPPTTAAACGEF----- 96
Db 638 CSPDERCPSTHFCHIGVGESENCCPKH--GDPQOALATGTGFSITRYTYDKETRRCRD 696
QY 97 -----YCPQGSDDCDIMDGPVCRDCKEKEPSTCASDGLTY 135
Db 697 FSYEGQGNANNTLTLEDGLVCPVLPNCSL--GEPPL-----SLQKEP-VICGGE--- 745
QY 136 YNRCYMAEACLGLHLHIVPCKHVLWPPSSPSPETTARPTPAAPVPPALVSSPSQ 195
Db 746 -----DTCPNGYVCHV-----GAPETT--NCCPCT-----RRPCL 775
QY 196 AVQVG--GTASL-----HCDVSGRPRAVYWEKOSHORENLIIMRPPDMYG 238
Db 776 PLEVGGVGEKLEKWFDDGIQMCPRPVYKGMKSNNTLT--KOSCRQSKENMPGCG--Y 832
QY 239 NVYVTSIGQVLT-----YNAREDAAGLYCTARNAAGLIRADPLSYVOREPARDAPS 292
Db 833 DPLVDTTGERMCTGQVRNPSQNS--YCHVGSALTTLTLC-----PRKIDP- 879
QY 293 IPAPAECLPDVOACTGPTSPHLYLMHDDPQGGCMTPFARGCDGAARGFETEACQ 350
Db 880 -----CQAVVEGTV--SEDLPRWFQKQNRCAPEYGVAGNENNFISQNTWCNACP 931
QY 351 -----CARGPQACVLPAYOGP 367
Db 932 EYRNYCPHGIPLIGSTVTSIGIDKGPBEFICMSSEFWNTICCDPMDFLSARDSCP 991
QY 368 CRGMEPRAAVSPLLQCGHPFYVYGCSENGNHNFSRSCEDACPVPRTPPCARCLRSKL 426
Db 992 CNNEKRGKRYDANTDTCEVYQYGGCEGTLINNFSLORCTEIC-----CKEYKRRLH 1043

```

RESULT 4

123573
 Hypothetical protein K10D3.4 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T33573
 R:McMurray, A.
 submitted to the EMBL Data Library, June 1996
 A:Reference number: 219762
 A:Accession: T33573
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-922 <MIL>
 A:Cross-references: EMBL:275545; PIDN:CAA9886.1; GSPDB:GN00019; CESP:K10D3.4
 A:Experimental source: clone K10D3
 C:Genetics:
 A:Gene: CESP:K10D3.4
 A:Map position: 1
 A:Introns: 60/1; 228/1; 278/1; 355/1; 743/1; 802/1; 885/2

Query Match 7.2% Score 216; DB 2; Length 922;
 Best Local Similarity 19.7% Pred. No. 1.5e-05;
 Matches 109; Conservative 45; Mismatches 166; Indels 234; Gaps 24;

```

Oy 31 GVCPLNOLSPNLMVD-----AOSTCERCSDRDCAAEKCCINVCGLHSCVAARF 80
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 GCPPLNGNPKFRINQVQOCYGADATCEAIGALSYDC-----ICSDDCCTVNN- 158
Oy 81 PGSPAAPTAAASC-----EGFVCPQOOGSDCDIMDQPV---CR-----CR 117
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 PISFCPSAFACIQPNNNGYTPPGGGTLLNMHYHDPITGEERELKYGGYGANNFQK 218
Oy 118 DRCEKEPSTCAGSDGLTYNN-----RCYDAEACGLGLHLHYPCNHYLSWPPS 166
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 219 DHCEYCKQKTC-NRGLPLRYRRTTGKQEPVYCCQGNNDGCG-----N 258
Oy 167 SPGPETTAAPRPGAAP-----VPPALYSPPQAVQVGSTASLHCDV-----S 210
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 259 NPNYCTTMTGLQCCPTTYLFTCSRRNGGIPSEVYNT-----ACGLPEYDVGIPDS 311
Oy 211 GNPRAVYTWKQSHORENLIMRPDMYGNVYTSIGQLVLYNARPDAGLYTCTARNAAG 270
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 312 GNTSPRFYVYDSRE-----GRCIQFS-----YLGCGGNFN 341
Oy 271 LIRADPPLSVVOR-----EPARDA-----ASIPRAPEC-----299
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 342 FLSQDHCEKFCFSRLCSAGEPLKDSGERNMECSPTGSGANCSPTHSCESTSGTTFGG 401
Oy 300 -----LPRVQACTGPTSPHLVLMHYDPQGGCTFPARCGDGAARGFETYEACQ 348
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 VCCPRPYVCKLPRQGNCGTYSNR--WTFNAKTGCEEFYISGCGANNFETYEKCO 458
Oy 349 QAC--ARG-----PGDA--358
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 DYCRRARSEPOCIQGTALTDNSGNFLICGSSASTPCANHYCYVDGTYGCCPTQAT 518
Oy 359 CYLPAVQV--PCRGWEPRAAYSPLLQCCHPFYVGGCEGNGNHNHRSCECDACVPRT 415
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 519 CSLSYKSGASGCPAVATRYWYSTRTCTQYTSFNGCDGNSNPNATQODCKDYRVESCPDG 578
Oy 416 -----PCRAC 420
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 579 GEVWKEONGNARAC 592

```

RESULT 5
 TIPGB1
 alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor - pig (fragment)
 N:Alternate names: b1kunitin; ITI; PI-14 (inhibitory fragment of ITI)
 C:Species: *Sus scrofa domestica* (domestic pig)
 C:Date: 30-Jun-1997 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: S11066; S13493; A01208
 R:Gebhard, W.; Schrettmueller, T.; Vetr, H.; Wachter, E.; Hochstrasser, K.

FEBS Lett. 269, 32-36, 1990
 A:Title: Complementary DNA and deduced amino acid sequences of porcine alpha1-microgl
 A:Reference number: S11066; MUID:90035395
 A:Accession: S11066
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-337 <GB>
 A:Cross-references: EMBL:X53685; NID:g1877; PIDN:CAA37725.1; PID:g1878
 R:Tavakkol, A.
 Biochim. Biophys. Acta 1088, 47-56, 1991
 A:Title: Molecular cloning of porcine alpha(1)-microglobulin/HI-30 reveals developmen
 A:Reference number: S13493; MUID:91113729
 A:Accession: S13493
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'M', 3-48, 'M', 50-337 <TAV>
 A:Cross-references: GB:X52087; NID:g1861; PIDN:CAA36306.1; PID:g1882
 A:Note: The authors translated the codon GTG for residue 2 as a Met initiation codon
 R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Reisinger, P.
 Biol. Chem. Hoppe-Seyler 366, 473-478, 1985
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte
 A:Reference number: A90685; MUID:85225967
 A:Accession: A01208
 A:Molecule type: protein
 A:Residues: 212-258, 'O', 260-269, 'S', 271-277, 'O', 279-282, 'A', 284, 'IR', 287-292, 'A', 294-
 C:Comment: This inhibitory fragment, released from native ITI after limited proteolys
 first domain interacts weakly with PMN-granulocytic elastase and not at all with panc
 C:Comment: The amino acid at position P2' (228-Met) appears to determine the specific
 nd elastase; those with leucine interact strongly.
 C:Superfamily: protein HC; animal kunitz-type proteinase inhibitor homology: 11pocal1
 C:Keywords: duplication; glycoprotein; plasma; serine proteinase inhibitor
 F:20-173/Domain: 11pocalin homology <IRP>
 F:216-266/Domain: animal kunitz-type proteinase inhibitor homology <BP1>
 F:217-322/Domain: animal kunitz-type proteinase inhibitor homology <BP2>
 F:216-266, 225-249, 241-262, 272-322, 281-305, 297-318/Disulfide bonds: #status predicted
 F:226/inhibitory site: Leu (chymotrypsin, elastase) #status predicted
 F:235/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:282/inhibitory site: Arg (trypsin) #status predicted

Query Match 7.1% Score 214; DB 1; Length 337;
 Best Local Similarity 39.0% Pred. No. 8.1e-06;
 Matches 41; Conservative 15; Mismatches 47; Indels 2; Gaps 1;

```

Oy 308 GPTSPHLVLMHYDPORGGCTFPARCCDGAARGFETYEACQQAARCPDACYLPAVQGP 367
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 223 GPCLGMIKRYFYNGSSMACETFPYGGCMGNGNFFVSEKECLQTCR--TVEACSLPIVSGP 280
Oy 368 CRGWEPRAAYSPLLQCCHPFYVGGCEGNGNHNHRSCECDACVP 412
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 281 CRGFQLMAFDVAVQGCVLNFYGGCGNGNPFSEKECKEYCGVP 325

```

RESULT 6
 S21089
 alpha-1-microglobulin/inter-alpha-trypsin inhibitor light chain precursor - rat
 N:Alternate names: acid-stable proteinase inhibitor; b1kunitin; trypstatin
 C:Species: *Rattus norvegicus* (Norway rat)
 C:Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
 C:Accession: S21089; A53056; A25935; A31890; A61633
 R:Lindqvist, A.; Bratt, T.; Altieri, M.; Kastern, W.; Akerstrom, B.
 Biochim. Biophys. Acta 1130, 63-67, 1992
 A:Title: Rat alpha(1)-microglobulin: co-expression in liver with the light chain of 1
 A:Reference number: S21089; MUID:92182014
 A:Accession: S21089
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-349 <LIN>
 A:Cross-references: GB:S87544; NID:g247162; PIDN:AA621782.1; PID:g247163
 R:Itton, H.; Ide, H.; Ishikawa, N.; Nawa, Y.
 J. Biol. Chem. 269, 3818-3822, 1994
 A:Title: Mast cell protease inhibitor, trypstatin, is a fragment of inter-alpha-tryps
 A:Reference number: A53056; MUID:94148692

A:Residues: 206-225 <EN62>
R:Taboni, C.; Cortese, R.
Nucleic Acids Res. 14, 6340, 1986
A:Title: Sequence of a full length cDNA coding for human protein HC (alpha-1-microglobulin)
A:Reference number: A25303; MUID:86312901
A:Accession: A25303
A:Molecule type: mRNA
A:Residues: 1-218, 'HW' <TRA>
A:Note: This mRNA sequence appears to contain errors after residue 218
R:Calero, M.; Escribano, J.; Grubb, A.; Mendez, E.
J. Biol. Chem. 269, 384-389, 1994
A:Title: Location of a novel type of interpeptide chain linkage in the human protein
A:Reference number: A53110; MUID:94103241
A:Accession: A53110
A:Molecule type: protein
A:Residues: 45-57 <CALI>
R:Vet, H.; Koegler, M.; Gebhard, W.
FEBS Lett. 245, 137-140, 1989
A:Title: The domain structure of the inhibitor subunit of human inter-alpha-trypsin inhibitor
A:Reference number: S03552; MUID:8917290
A:Accession: S03552
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 206-352 <VER2>
R:Malik, N.; Baldyck, M.; Maes, P.; Capon, C.; Han, K.K.; Tartar, A.; Fournier
Biochim. Biophys. Acta 1009, 1009-1018, 1992
A:Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolation
A:Reference number: S28928; MUID:93039735
A:Accession: S28928
A:Status: preliminary
A:Molecule type: protein
A:Residues: 206-215 <MAL>
R:Morelle, W.; Capon, C.; Baldyck, M.; Sautiere, P.; Kouach, M.; Michalski, C.; Fournier
Eur. J. Biochem. 221, 881-888, 1994
A:Title: Chondroitin sulphate covalently cross-links the three polypeptide chains of inter-alpha-trypsin inhibitor
A:Reference number: S43466; MUID:94229087
A:Accession: S43466
A:Status: preliminary
A:Molecule type: protein
A:Residues: 206-221 <WOR>
R:Wislowski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Valcek, J.
Biochemistry 33, 7423-7429, 1994
A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable complex
A:Reference number: A53642; MUID:94271799
A:Accession: A53642
A:Status: preliminary
A:Molecule type: protein
A:Residues: 206-217 <WIS>
R:Calero, M.; Mendez, E.; Garcia, E.
Biochim. Biophys. Acta 1149, 91-99, 1995
A:Title: Expression of the human complex-forming glycoprotein HC (alpha-1-microglobulin)
A:Reference number: S55688; MUID:95284116
A:Accession: S55688
A:Molecule type: protein
A:Residues: 20-24 <CAL2>
R:Bourguignon, J.; Diarra-Mehrpour, M.; Sesboue, R.; Fraai, M.; Sala-Trepat, J.M.; Martini
Biochem. Biophys. Res. Commun. 131, 1146-1153, 1985
A:Title: Human inter-alpha-trypsin-inhibitor: characterization and partial nucleotide sequence
A:Reference number: I52208; MUID:86025577
A:Accession: I52208
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 302-352 <BOU>
A:CROSS-references: GB:M11552; NID:g166587; PIDN:AAA59194.1; PID:g307077
R:Wojcik, E.G.C.; van den Berg, M.; van der Linden, I.K.; Poort, S.R.; Cupers, R.; Berth
Biochem. J. 311, 753-759, 1995
A:Title: Factor IX zuphan: a Cys(18) -> Arg mutation results in formation of a heterodimer
A:Reference number: S59509; MUID:96067589
A:Accession: S59509
A:Molecule type: protein
A:Residues: 27-35, 'Y', 37 <WOJ>
R:Altman, F.; Mizon, J.; Khan, S.R.
Eur. J. Biochem. 236, 984-990, 1996

A>Title: Identification of uronic-acid-rich protein as urinary bikunin, the light chain of alpha-1-macroglobulin.
A:Reference number: S66434; MUID:96270753
A:Accession: S66434
A:Molecule type: Protein
A:Residues: 206-214, 'X', 216-230 <ATM2>
R:Akerstrom, B.; Bratt, T.; Englund, J.J.
FEBS Lett. 362, 50-54, 1995
A>Title: Formation of the alpha(1)-microglobulin chromophore in mammalian and insect hemoglobin.
A:Reference number: S68728; MUID:95212582
A:Accession: S68728
A:Molecule type: Protein
A:Residues: 89-100 <AKE>
R:Jessen, T.E.; Faarvang, K.L.; Ploug, M.
FEBS Lett. 230, 195-200, 1988
A>Title: Carbohydrate as covalent crosslink in human inter-alpha-trypsin inhibitor: a new family of proteins.
A:Reference number: S02431; MUID:86167187
A:Accession: S02431
A:Molecule type: Protein
A:Residues: 206-214, 'X', 216-217 <RES>
R:Lopez, C.; Grubb, A.; Mendez, E.
FEBS Lett. 144, 349-353, 1982
A>Title: Human protein HC displays variability in its carboxyl-terminal amino acid sequence.
A:Reference number: A91304
A:Contents: annotation; Variant of alpha-1-microglobulin
A>Note: pooled urine samples contained two forms of this protein, both lacking 57-Tyr
R:Hochstrasser, K.; Schonberger, O.L.; Rossmannith, I.; Wächter, E.
Hoppe-Sevler's Z. Physiol. Chem. 362, 1357-1362, 1981
A>Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-alpha-trypsin inhibitor.
A:Reference number: A91698; MUID:82074265
A:Contents: annotation; carbohydrate binding sites
R:Morill, M.; Travis, J.
Biol. Chem. Hoppe-Sevler 366, 19-21, 1985
A>Title: The reactive site of human inter-alpha-trypsin inhibitor is in the amino-terminal part of the molecule.
A:Reference number: A90682; MUID:85225940
A:Contents: annotation; inhibitory site
A>Note: in vitro, the first twelve residues of the amino end of the inhibitor appear to be identical to those of the inter-alpha-trypsin inhibitor.
C:Comment: Alpha-1-microglobulin and inter-alpha-trypsin inhibitor are proteolytically cleaved at the same site in many physiological fluids including plasma . It contains at least one brown-yellow chromophore.

Query Match 7.1%; Score 213; DB 1; Length 352;
Best Local Similarity 40.2%; Pred. No. 9.7e-06;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

QY 306 CGGPSPLVLMHYDPORGCGMTFPAAGCDGARGEFERYEACDAGCARGPDCAVLPAYQ 365
| | | | : | : | | | | | | | | | | | |
Db 240 CNGMTSRFF----YGTSMACETFYGGCMGNNGNFTYEKELQTGRVA--ACNLPIVR 293
| | | | : | : | | | | | | | | | | | |

QY 366 GPCRGMEPRMAYSPLLQOCHPEFYVGCEGCNGNFHSRESCEDACPVP 412
| | | | : | : | | | | | | | | | | | |
Db 294 GCRRFAFLTMADFADVKAKCVLFPPYGCGQGNGNKFSSEKECEYCGVP 340
| | | | : | : | | | | | | | | | | | |

RESULT 8
TTBOBI
alpha-1-microglobulin / inter-alpha-trypsin inhibitor precursor [validated] - bovine
N:Alternate names: BI-14 (inhibitory fragment of ITI); bikunin; ITI
C:Species: Bos primigenius taurus (cattle)
C>Date: 25-Feb-1965 #sequence revision 04-Feb-2000 #text-change 18-Aug-2000
C:Accession: S68149; A91717; A90685; S31219; A01209
R:Lindqvist, A.; Akerstrom, B.
Biochim. Biophys. Acta 1306, 98-106, 1996
A>Title: Bovine alpha(1)-microglobulin/bikunin. Isolation and characterization of live
A:Reference number: S68149; MUID:96201710
A:Accession: S68149
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-352 <LIN>
A:Cross-references: EMBL:U35642; NID:g1016297; PIDN:AMB07599.1; PID:g1016298
R:Hochstrasser, K.; Wächter, E.
Hoppe-Sevler's Z. Physiol. Chem. 364, 1679-1687, 1983

```

A:Residues: 3-125 <HOC>
R:Veerataavan, K.; Singh, K.; Wachter, E.; Hochstrasser, K.
Biochem. Int. 26, 405-413, 1992
A:Title: Characterization of a trypsin inhibitor from equine urine.
A:Reference number: A45653; MUID:92328813
A:Accession: A45653
A:Status: preliminary
A:Molecule type: Protein
A:Residues: 1-12, 'E', 14-33 <VER>
A:Cross-references: PIDD:AA82430.1; PID:g250686
A:Experimental source: urine
A>Note: sequence extracted from NCBI backbone (NCBIP:107966)
C:Comment: This inhibitory fragment, released from native IPI after limited proteolysis, first domain interacts weakly with PN-7-granulocytic elastase and not at all with pancreatic elastase; The amino acid at position p2' (19-Met) appears to determine the specificity.
C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology <BP1>
C:Keywords: duplication; glycoprotein; plasma; serine proteinase inhibitor
F:63-113/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
F:7-57/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
F:7-57,16-40,32-53,63-113,72-96,88-109/Disulfide bonds: #status predicted
F:17/inhibitory site: Ieu (chymotrypsin, elastase) #status predicted
F:26/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:73/inhibitory site: Arg (trypsin) #status predicted

Query Match 6.9%; Score 208; DB 1; Length 125;
Best Local Similarity 38.1%, Pred. No. 7,6e-06;
Matches 40; Conservative 14; Mismatches 49; Indels 2; Gaps 1;

QY 308 GPTSPHLVLMHYDPQRGGCTFPARGCDGAAGFETFEACQACARGPDACVLPAYOGP 367
    || : : : : || || || : : : : || || ||
Db 14 GPCLGMSIRFNYNGTSMACETFYGGCLGNGNNFASQKECLOQTCTVA--ACNLPYOGP 71

QY 368 CRGWEPRWVAYSPLLQOCHPFVYGGCGGNGNNFHSRSCEDACVPV 412
    || : : || : || || || || || || : : : : ||
Db 72 CRAFIRLMAFDAAGKCVLFTYGGCGNGNKFYSQKEKEYCGIP 116

RESULT 10
A29652
Inter.alpha.-trypsin inhibitor (BPI type) - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #textL_change 16-Jul-1999
C:Accession: A29652
R:Asp, G.; Hochstrasser, K.; Wachter, E.; Reisinger, P.W.M.
Biol. Chem. Hoppe-Seyler 368, 727-731, 1987
A:Title: The amino-acid sequence of the trypsin-released inhibitor from sheep inter-a
sin inhibitor, XI.).
A:Reference number: A29652; MUID:87299012
A:Accession: A29652
A:Molecule type: protein
A:Residues: 1-123 <RAS>
C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocali
C:Keywords: serine proteinase inhibitor
F:55/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
F:61-111/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>

Query Match 6.8%; Score 205; DB 2; Length 123;
Best Local Similarity 37.1%, Pred. No. 1,1e-05;
Matches 39; Conservative 15; Mismatches 49; Indels 2; Gaps 1;

QY 308 GPTSPHLVLMHYDPQRGGCTFPARGCDGAARCFEYEAQCAARGPDACVLPAYOGP 367
    || : : : : || || || : : : : || || ||
Db 12 GPCLGMSIRFNYNGTSMACETFYGGCGGNGNNFBSRSCEDACVPV 69

QY 368 CRGWEPRWVAYSPLLQOCHPFVYGGCGGNGNNFHSRSCEDACVPV 412
    || : : || : || || || || || || : : : : ||
Db 70 CRAFIRLMAFDAAGKCVLFTYGGCGNGNKFYSQKEKEYCGIP 114

```

Query Match	6.78	Score	203.5	DB	2	Length	299
Best Local Similarity	31.2%	Pred. No.	3e-05				
Matches	50	Conservative	20	Indels	71	Gaps	19
QY	270	GLLADPPLSVQKEP----	ARDAAPSIPAPAECLDPVQACTGETSPHLVLMHYDPORG	325			

A:Status: preliminary; translated from GR/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1743 <NLT>
A:CROSS-references: EMBL:AL032623; PIDN:CAA2151.1; CESP:V43FB8.3
A:Experimental source: clone V43FB8
C:Genetics:

A1Gene: CESP:Y43F8B.3
A1Introns: 65/1; 92/2; 128/1; 229/1; 367/1; 422/1; 486/2; 533/1; 571/1; 628/1; 657/2; 694/1

Query Match

Best Local Similarity	6.7%;	Score 202.5;	DB 2;	Length 1743
Matches 124:	Conservative	22.18;	Pred. No. 0.00017;	
	Mismatch	49.		

```

QY      52 EGSDDCAAHEKCCINWGLSHCSVAARPPGSPAAP-----TTAASCE   94
Db      1019 :::|||OVCPEGHYCHVTEANTYCCKA--LGSPCCGGPLRDVGWGSGLSWMYNQSSQCCL 1076
QY      95 GF-VCPQGGS-----DODIWDGPVCRRCRCKEKESFTICASDGLTYVR--CYND 142
Db      1077 PFSYGCGKGTQNUNFLTKQDCD-----RTCYELDNPCALGDPMQANNRRPLOCSAT 1126
QY      143 AEAACLGLHLHIIVCKHVLWMPSSPGPETTAPT-----PGAAPVPALYSPPQAQV 199
Db      1127 ASTCGAQFCHFGANGDPTVCCPGRVESPQIQOGMAVGTGATLPFRWTYNQTQOCYOF 1186
QY      200 GGTAISHCVSVSRPPPAVTWEKOSHORENLIMRPDMYTGNVVTSIGOLVIYNARPDAAG 259
Db      1187 -----NYAAR-----MGNNNFOSQAOECBOTPYVNVNCPGTSPLHDASTNKP----- 1229
QY      260 LYTC-ARNAAGLLRADF--PLSYVQRE-PARDAAPSTIPAARECLPDVOACTGPTSPHV 315
Db      1230 -VPCTFGNSISCG--ADMHCHLGVLPDEYOCCPGSFTNPBGACOOGLPESEGEVTGAAPAPPTS 1285
QY      316 LMHTYPORGKCTEPFARGCDDGAARGFETYEACQOAC-----ARGPDA 358
Db      1286 RMYUDOTMCKOEFTYNGRGRNONNFLUQEDCAATCDVFTNPCOMIALPATLTCSGGISS 1345
QY      359 -----CVLPAYOGP---CRGWEPRAVSPYLQQCHEPVYGGCEGNGNHFHSRESCD 407
Db      1346 DTGCANNMCHIGAANDSTVCC-----PSGRKRKNFOOCLFEQYNKRGQNNFNENQADOER 1401
QY      408 ACVPVPRTP-----PCRACRLRSKLATSLTRS--DLAIYGRLE----- 443
Db      1402 TCPPIFINPCLGLEVITLEDGAPRKCK-----PLLNSSGSAETEFCHTNPDPDONSFFCCPR 1455
QY      444 VLEBP-----EAGAIGIAVLAEDVLDK--DKMKGL-----PLGTGYLVLEVTLSGMD 486
Db      1456 INDDPCNAFVANRNGNFMNTRYYNPVEGDOFSFYHGILNGENNNFLTLMKCDETCKPFLT 1515
QY      487 WACPCCMMTAGDGFLVMIGFY 507
Db      1516 TAC-----FGEGSEFLMNNGRV 1531

```

RESULT 15

hypothetical protein ZC84.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*

C:\species: Caenorhabditis elegans

```
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Aug-1997
C;Accession: 320201
```

C;Accession: S28291

R; Thomas, K.

submitted to the EMBL Data Library, December 1992.

A; Reference number: 528285

A;Accession: S28291

```

//SESSION: 328291
A;Status: Preliminary

```

A;Molecule type: DN

A: molecule type: DNA
A: Residues: 1-2844 <THO>

A;Residues: 1-2844 <THO>

A; Cross-re:

A; Introns: 14/1; 32/3; 57/1; 192/3; 277/1; 398/1; 439/1; 474/1; 497/1; 813/1; 1135/1; 1249/3/1; 2555/1; 2720/1; 2739/3; 2819/1
...Superfamily)''

C:Superfamily: animal kunitz-type proteinase inhibitor homolog

EC:243-205-2000-274/Domain: animal kunitz-type proteinase inhibitor homolog

F;343-395/Domain: animal Kunitz-type proteinase inhibitor homolog

F;442-492/Domain: animal kunitz-type proteinase inhibitor homol

F;546-598/Domain: animal Kunitz-type proteinase inhibitor homol

[illegible]

EF1662-1776/Domain: animal kunitz-type proteinase inhibitor homology <BPI5>

F_1787-1833/Domain:	animal	Kunitz-type proteinase inhibitor	homology	<BP17>
F_1845-1893/Domain:	animal	Kunitz-type proteinase inhibitor	homology	<BP18>
F_1952-2004/Domain:	animal	Kunitz-type proteinase inhibitor	homology	<BP19>
F_2097-2152/Domain:	animal	Kunitz-type proteinase inhibitor	homology	<BP20>

Query Match	6.6%;	Score 200.5;	DB 2;	Length 2844
Best Local Similarity	21.9%;	Prod No 0.00024		

[illegible]

Search completed: February 26, 2002, 01:16:14
Job time: 534 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 26, 2002, 01:20:11 ; Search time 29.31 Seconds

(without alignments)
2734.808 Million cell updates/sec

Title: US-09-819-136-2

Perfect score: 548
Sequence: 1 MPALRPLPLLLRLTSGA.....KRLLELLEKQACELNRFQD 548

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size: 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	1.8	58	5	09TXD3
2	10	1.8	154	6	09NOX3
3	10	1.8	169	6	09NOX7
4	9	1.6	18	4	09UM92
5	9	1.6	251	4	09UP15
6	9	1.6	281	4	015845
7	9	1.6	290	4	012876
8	9	1.6	290	4	09UD69
9	9	1.6	301	4	09BW70
10	9	1.6	311	11	09QUL7
11	9	1.6	321	4	09UBA2
12	9	1.6	321	4	09NRB2
13	9	1.6	321	4	09NRB2
14	9	1.6	322	4	09NRO8
15	9	1.6	335	6	09NYC8
16	9	1.6	335	6	09GK80
17	9	1.6	351	6	09NOX8
18	9	1.6	422	4	075462
19	9	1.6	422	4	09UHH5
19	9	1.6	829	4	09H684

20	9	1.6	1216	11	062884	062884 ratu
21	9	1.6	1342	4	09H422	09H422 homo sapien
22	9	1.6	1549	11	060444	060444 cricetus
23	9	1.6	1884	4	09ULD7	09ULD7 homo sapien
24	9	1.6	2035	4	075035	075035 homo sapien
25	9	1.6	2920	11	09ROM0	09ROM0 mus musculu
26	9	1.6	2944	11	063870	063870 mus musculu
27	9	1.6	3198	5	09H868	09H868 manduca sex
28	8	1.5	58	9	09A273	09A273 bacterioph
29	8	1.5	105	5	09NMV2	09NMV2 leishmania
30	8	1.5	106	10	041051	041051 pism saliv
31	8	1.5	122	2	067176	067176 aquifex aeo
32	8	1.5	143	4	015412	015412 homo sapien
33	8	1.5	147	11	09WUP1	09WUP1 mus musculu
34	8	1.5	147	11	09JMB8	09JMB8 ratu
35	8	1.5	147	11	09JMB8	09JMB8 ratu
36	8	1.5	185	10	09LUV0	09LUV0 arabidopsi
37	8	1.5	187	5	09NDF1	09NDF1 heterodera
38	8	1.5	227	4	09UKJ0	09UKJ0 homo sapien
39	8	1.5	231	11	09ERM0	09ERM0 mus musculu
40	8	1.5	245	11	09JMG3	09JMG3 mus musculu
41	8	1.5	246	4	09AVT8	09AVT8 homo sapien
42	8	1.5	264	2	09PC52	09PC52 xyella fas
43	8	1.5	267	2	09CLU6	09CLU6 pasteurella
44	8	1.5	271	4	09HBS0	09HBS0 homo sapien
45	8	1.5	272	11	070356	070356 mus musculu

ALIGNMENTS

RESULT 1
ID Q9TXD3 PRELIMINARY: PRT: 58 AA.
AC Q9TXD3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FH-KTM-6.751 KDA MONOMERIC KUNITZ-TYPE PROTEINASE INHIBITOR.
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Platyhelminthes; Echinostomida; Fasciolidae;
OC Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE.
RX MEDLINE=96360459; PubMed=8719242;
RA Bozas S.E., Panaccio M., Creaney J., Dosen M., Parsons J.C.,
RA Vlasuk G.V., Walker I.D., Spithill T.W.;
RT "Characterisation of a novel kunitz-type molecule from the trematode
RT Fasciola hepatica."
RL Mol. Biochem. Parasitol. 74:19-29(1995).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR HSSP: P02760; IBIK.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00131; KU; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor.
SQ SEQUENCE 58 AA: 6589 MW: 28741CA4B0757AD3 CRC64;

Query Match 1.8%; Score 10; DB 5; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 386 PVTYGGCEGN 395
DB 32 PVTYGGCEGN 41

```

RESULT 2
ID 09NOX3 PRELIMINARY; PRT; 154 AA.
AC 09NOX3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KUNITZ DOMAIN PROTEIN 2 (FRAGMENT).
GN TKD2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PERI-IMPLANTATION WHOLE CONCEPTUS;
RA Green J.A., Maclean J.A. II, Gan X., Xie S., Roberts R.M.;
RT "The cloning and analysis of expression of trophoblast Kunitz domain
RT proteins (TKDP).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AF241828; AAF64495.1; -
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00131; KU: 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR Serine protease inhibitor.
FT NON_TER 1
SQ SEQUENCE 154 AA; 17069 MW; B81A91E83212F52A CRC64;

Query Match
Best local Similarity 1.8%; Score 10; DB 6; Length 154;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 PFYGGCEGN 395
DB 124 PFYGGCEGN 133

RESULT 3
ID 09NOX7 PRELIMINARY; PRT; 169 AA.
AC 09NOX7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TROPHOBLAST KUNITZ DOMAIN PROTEIN 2.
GN TKDP1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BREED STIMENTAL X HEREFORD;
RA Green J.A., Maclean J.A. II, Gan X., Xie S., Roberts R.M.;
RT "Cloning and characterization of the expression of trophoblast Kunitz
RT domain proteins (TKDP).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AF241777; AAF61248.1; -
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00131; KU: 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR Serine protease inhibitor.
FT
SQ

```

```

SQ SEQUENCE 169 AA; 18855 MW; B65C16F4DF0BA3D CRC64;

Query Match
Best local Similarity 1.8%; Score 10; DB 6; Length 169;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 PFYGGCEGN 395
DB 139 PFYGGCEGN 148

RESULT 4
ID 09UM92 PRELIMINARY; PRT; 18 AA.
AC 09UM92;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE UPAR PROTEIN (FRAGMENT).
GN UPAR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95329719; PubMed=7605992;
RA Soravia E., Grebe A., De Luca F., Helin K., Suh T.T., Degen J.L.,
RA Blas F.;
RT "A conserved TATA-less proximal promoter drives basal transcription
RT from the urokinase-type plasminogen activator receptor gene.";
RL EMBL; S78532; AAD14289.1; -
FT NON_TER 18
SQ SEQUENCE 18 AA; 1964 MW; 617216DF83DE726C CRC64;

Query Match
Best local Similarity 1.6%; Score 9; DB 4; Length 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PLPLPLLL 14
DB 5 PLPLPLLL 13

RESULT 5
ID 09UP15 PRELIMINARY; PRT; 251 AA.
AC 09UP15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE U-PLASMINOGEN ACTIVATOR RECEPTOR FORM 2 PRECURSOR-HUMAN (FRAGMENT).
GN PLAUR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Lameudin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attik C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce K., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Garrano A.V.;
RT "Sequence analysis of a 4.2 kb region in 19q13.2 between AKT2 and
RT D19S178.";
SQ Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

```


DR EMBL: AC006953; AAD17389.1; -
 DR InterPro: IPR001526; LY6_UPAR.
 DR Pfam: PF00021; UPAR_LY6_2.
 DR PROSITE: PS00983; LY6_UPAR; 2.
 DR SMART: SM00134; LU; 2.
 KW Receptor.
 FT NON_TER 251
 SQ SEQUENCE 251 AA; 27923 MW; 17285C5B26B3BA4D CRC64;

Query Match 1.6%; Score 9; DB 4; Length 251;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PLPLPLLL 14
 |||||
 DB 5 PLPLPLLL 13

RESULT 6

ID 015845 PRELIMINARY; PRT; 281 AA.

AC 015845;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UROKINASE PLASMINOGEN ACTIVATOR RECEPTOR PRECURSOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE-9314820; PubMed-8392005;
 RA Pyke C., Eriksen J., Solberg H., Schnack Nielsen B., Kristensen P.,
 RA Lund L.R., Danoe K.;
 RT "An alternatively spliced variant of mRNA for the human receptor for
 RT urokinase plasminogen activator."
 RL FEBS Lett. 326:69-74(1993).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Pyke C.;
 RL Submitted (NOV-1993) to the EMBL/Genbank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Pyke C.;
 RL Submitted (FEB-1994) to the EMBL/Genbank/DBJ databases.
 DR EMBL: X74039; CAA52191.1; -
 DR InterPro: IPR001526; LY6_UPAR.
 DR Pfam: PF00021; UPAR_LY6_2.
 DR PROSITE: PS00983; LY6_UPAR; 3.
 DR SMART: SM00134; LU; 2.
 KW Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 278 UROKINASE PLASMINOGEN ACTIVATOR
 FT RECEPTOR.

FT CHAIN 23 278 UROKINASE PLASMINOGEN ACTIVATOR
 FT RECEPTOR.
 SQ SEQUENCE 281 AA; 31263 MW; E6C580F279FB0316 CRC64;

Query Match 1.6%; Score 9; DB 4; Length 281;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PLPLPLLL 14
 |||||
 DB 5 PLPLPLLL 13

RESULT 7

ID 012876 PRELIMINARY; PRT; 290 AA.

AC 012876;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UROKINASE-TYPE PLASMINOGEN ACTIVATOR RECEPTOR.
 GN PLAU.R.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Casey J.R., Petranka J.G., Kottra J., Fleenor D.E., Rosse W.F.;
 RL Submitted (APR-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE-90151622; PubMed-1689240;
 RA Roldan A.L., Cubellis M.V., Masucci M.T., Behrendt N., Lund L.R.,
 RA Dano K., Appella E., Blasi F.;
 RT "Cloning and expression of the receptor for human urokinase
 RT plasminogen activator, a central molecule in cell surface, plasmin
 RT dependent proteolysis."
 RL EMBO J. 9:467-474(1990).
 RN [3]

RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
 RA Burkhart-Schultz K., Gordon L., Dias J., Ramirez M., Stllwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
 RA Dangnanan L., Ertler A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Attix C., Andreise T., Tranckheim W., Amico-Keller G., Coelield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of a 4.2 Mb region in 19q13.2 between AKT2 and
 RT D19S178."
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U08839; AAB60333.1; -
 DR InterPro: IPR001526; LY6_UPAR.

DR Pfam: PF00021; UPAR_LY6_2.
 DR PROSITE: PS00983; LY6_UPAR; 2.
 DR SMART: SM00134; LU; 2.
 KW Receptor; Kinase.
 SQ SEQUENCE 290 AA; 32016 MW; 60F14E2A2AFB67AC CRC64;

Query Match 1.6%; Score 9; DB 4; Length 290;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PLPLPLLL 14
 |||||
 DB 5 PLPLPLLL 13

RESULT 8

ID 09UD69 PRELIMINARY; PRT; 290 AA.

AC 09UD69;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UROKINASE RECEPTOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE-94178507; PubMed-8131971;
 RA Bayraktutan U., Jones P.;
 RT "A novel urokinase receptor on monocyte-like macrophage cell line."
 RL Biochem. Soc. Trans. 21:395-395(1993).
 DR InterPro: IPR001526; LY6_UPAR.
 DR Pfam: PF00021; UPAR_LY6_2.

DR PROSITE; PS00983; LY6_UPAR; 2.
 DR SMART; SM00134; L0; 2.
 SQ SEQUENCE 290 AA; 32030 MW; 60F1568920C8753C CRC64;

Query Match
 Best Local Similarity 1.6%; Score 9; DB 4; Length 290;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 PLPLPLLLL 14
 |||||
 Db 5 PLPLPLLLL 13

RESULT 9
 Q9BMT0 PRELIMINARY; PRT; 301 AA.
 AC Q9BMT0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 GN SOLUBLE UROKINASE PLASMINOGEN ACTIVATOR RECEPTOR PRECURSOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fu J., Bal X., Wang M., Xi X., Ruan C.;
 RT "Cloning and expression of the soluble receptor for human urokinase
 RT plasminogen activator.";
 RL Chung Hua Hsueh yeh Hsueh Tsa Chih 22:0-0(2001).
 DR EMBL; AY029180; AAK31795.1; -.
 KW Signal; Receptor; Kinase.
 FT SIGNAL 1 22 POTENTIAL.
 SQ SEQUENCE 301 AA; 33304 MW; 98E264C1D18D988 CRC64;

Query Match
 Best Local Similarity 1.6%; Score 9; DB 4; Length 301;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 PLPLPLLLL 14
 |||||
 Db 5 PLPLPLLLL 13

RESULT 10
 Q9QUL7 PRELIMINARY; PRT; 311 AA.
 AC Q9QUL7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 GN TRANSMEMBRANE TRYPTASE.
 GN TP5G1 OR TMT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129SV, AND BALB/C;
 RX MEDLINE=9452974; PubMed=10521469;
 RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,
 RA Friend D.S., Krilis S.A., Stevens R.L.;
 RT "Identification of a New Member of the Trypsin Family of Mouse and
 RT Human Mast Cell Proteases Which Possesses a Novel COOH-terminal
 RT Hydrophobic Extension.";
 RL J. Biol. Chem. 274:30784-30793(1999).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).

DR EMBL; AF175760; AAF03698.1; -.
 DR EMBL; AF175523; AAF03696.1; -.
 DR HSSP; P20231; IAAO.
 DR MGD; MGI:1349391; Tpsg1.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PF00089; trypsin.1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP-SPC; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN.1.
 KW Hydrolase; Serine protease; Transmembrane.
 SQ SEQUENCE 311 AA; 32656 MW; 7FC9D6E6A2A86808 CRC64;

Query Match
 Best Local Similarity 1.6%; Score 9; DB 11; Length 311;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 351 CARGPGDAC 359
 |||||
 Db 202 CARGPGDAC 210

RESULT 11
 Q9UBB2 PRELIMINARY; PRT; 321 AA.
 AC Q9UBB2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 GN TRANSMEMBRANE TRYPTASE.
 GN TMT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9452974; PubMed=10521469;
 RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,
 RA Friend D.S., Krilis S.A., Stevens R.L.;
 RT "Identification of a New Member of the Trypsin Family of Mouse and
 RT Human Mast Cell Proteases Which Possesses a Novel COOH-terminal
 RT Hydrophobic Extension.";
 RL J. Biol. Chem. 274:30784-30793(1999).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 CC EMBL; AF175759; AAF03697.1; -.
 DR EMBL; AF175522; AAF03695.1; -.
 DR HSSP; P00763; IDPO.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin.1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP-SPC; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN.1.
 KW Hydrolase; Serine protease; Transmembrane.
 SQ SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4962D CRC64;

Query Match
 Best Local Similarity 1.6%; Score 9; DB 4; Length 321;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 351 CARGPGDAC 359
 |||||
 Db 210 CARGPGDAC 218

RESULT 12
 Q9NRR2 PRELIMINARY; PRT; 321 AA.
 ID Q9NRR2
 AC Q9NRR2;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE GAMMA I-TRYPTASE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20302813; PubMed=10843716;
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaro M.,
 RA Wolters P.J., Vergese G.M.;
 RT "Characterization of Human gamma-tryptases, Novel Members of the
 RT Chromosome 16p Mast Cell Tryptase and Proctasin Gene Families";
 RL J. Immunol. 164:6566-6575(2000).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL: AF191031; AAF76457.1; -.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; Tryp_Spc.1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 321 AA; 33727 MW; 1AE00D160420562B CRC64;

Query Match 1.6%; Score 9; DB 4; Length 321;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 CARGPDAC 359
 DB 210 CARGPDAC 218

RESULT 13
 ID 09NR08 PRELIMINARY; PRT; 321 AA.
 AC 09NR08;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE GAMMA II-TRYPTASE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20302813; PubMed=10843716;
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaro M.,
 RA Wolters P.J., Vergese G.M.;
 RT "Characterization of Human gamma-tryptases, Novel Members of the
 RT Chromosome 16p Mast Cell Tryptase and Proctasin Gene Families";
 RL J. Immunol. 164:6566-6575(2000).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL: AF195508; AAF76458.1; -.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; Tryp_Spc.1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 321 AA; 33861 MW; 1EF8C29C0A9CDAA2 CRC64;

Query Match 1.6%; Score 9; DB 4; Length 321;
 Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 CARGPDAC 359
 DB 210 CARGPDAC 218

RESULT 14
 ID 09NYC8 PRELIMINARY; PRT; 322 AA.
 AC 09NYC8;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UROKINASE-TYPE PLASMINOGEN ACTIVATOR RECEPTOR (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GIANT CELL LUNG CANCER;
 RA Zhu F., Jia S., He F.;
 RT "Molecular cloning of UPAR from a human giant cell lung cancer cell
 RT line";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF257789; AAF71751.1; -.
 DR InterPro: IPR001526; LY6_UPAR.
 DR Pfam: PF00021; UPAR_LY6; 2.
 DR SMART: SM00134; LY6; 3.
 DR PROSITE: PS00983; LY6_UPAR; 3.
 KW Receptor; Kinase.
 FT NON_TER 322
 SQ SEQUENCE 322 AA; 35476 MW; 7E9CEDD0162950FD CRC64;

Query Match 1.6%; Score 9; DB 4; Length 322;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PLPLPLLL 14
 DB 5 PLPLPLLL 13

RESULT 15
 ID 09GR80 PRELIMINARY; PRT; 335 AA.
 AC 09GR80;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UROKINASE PLASMINOGEN ACTIVATOR RECEPTOR.
 GN UPAR.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKIN;
 RA Engelholm L.H., Behrendt N.;
 RT "Characterization of the urokinase plasminogen activator receptor on
 RT primate cell lines";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF302071; AAG40759.1; -.
 DR InterPro: IPR001526; LY6_UPAR.
 DR InterPro: IPR003631; U-PAR.
 DR Pfam: PF00021; UPAR_LY6; 2.
 DR ProDom: PD007582; U-PAR; 1.
 DR SMART: SM00134; LY6; 3.
 DR PROSITE: PS00983; LY6_UPAR; 1.
 KW Receptor; Kinase.

SO SEQUENCE 335 AA: 37022 MW: C56F3A1621BA2D64 CRC64;

Query Match 1.6%; Score 9; DB 6; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PLPLPLLL 14
|||
Db 5 PLPLPLLL 13

Search completed: February 26, 2002, 01:23:46
Job time: 215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:17:55 ; Search time 14.59 Seconds
(without alignments)
1377.130 Million cell updates/sec

Title: US-09-819-136-2

Sequence: 1 MPALRPPLPPLLLLRUTSGA.....KKILELLEKQACELLNRFQD 548

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	2.4	252	1 SPT2_MOUSE	Q9WU03 mus musculus
2	10	2.4	252	2 EFA2_MOUSE	P52801 mus musculus
3	10	1.8	213	1 EFA2_HUMAN	Q43921 homo sapien
4	9	1.6	164	1 TKD1_BOVIN	Q28201 bos taurus
5	9	1.6	265	1 TKD1_SHEEP	Q29428 ovis aries
6	9	1.6	328	1 IBP2_HUMAN	P18065 homo sapien
7	9	1.6	335	1 UPAR_HUMAN	Q03405 homo sapien
8	9	1.6	530	1 PVR2_MOUSE	P32507 mus musculus
9	9	1.6	1053	1 ITRAS_MOUSE	P16888 mus musculus
10	9	1.6	1238	1 PRPJ_MOUSE	Q64455 mus musculus
11	9	1.6	2944	1 CA17_HUMAN	Q02388 homo sapien
12	8	1.5	27	1 ACH4_MOUSE	Q70174 mus musculus
13	8	1.5	100	1 APC2_CAVPO	P27916 cavia porce
14	8	1.5	122	1 YA84_AQUAE	Q67176 aquifex ae
15	8	1.5	148	1 RMP3_HUMAN	Q60896 homo sapien
16	8	1.5	184	1 MUPM_MOUSE	P04939 mus musculus
17	8	1.5	249	1 YGRS_STRCO	P40179 streptomyce
18	8	1.5	281	1 NAE1_THEMA	Q9WZB3 thermotoga
19	8	1.5	461	1 NCBI_HUMAN	Q02818 homo sapien
20	8	1.5	487	1 MORT5_HUMAN	O15374 homo sapien
21	8	1.5	527	1 G19P_HUMAN	P14314 homo sapien
22	8	1.5	630	1 ACH4_RAT	P09483 rattus norv
23	8	1.5	650	1 APPL_HUMAN	P51693 homo sapien
24	8	1.5	662	1 MOC1_XENLA	Q05049 xenopus lae
25	8	1.5	699	1 PTPE_MOUSE	P23469 mus musculus
26	8	1.5	700	1 GP85_TRYCR	Q03877 trypanosoma
27	8	1.5	714	1 CDG2_HUMAN	Q09514 homo sapien
28	8	1.5	1007	1 NEOL_MOUSE	P97198 mus musculus
29	8	1.5	1493	1 Y137_TREPA	O83173 treponema p
30	7	1.3	45	1 ITR4_RADMA	P16344 radianthus
31	7	1.3	56	1 FER_METTE	Q01700 methanosarc
32	7	1.3	58	1 CD1B_RABIT	P23042 oryctolagus
33	7	1.3	109	1	

ALIGNMENTS

RESULT	ID	SPRT2_MOUSE	STANDARD	PRT	252 AA
34	7	1.3	120	1	SYZ7_MOUSE
35	7	1.3	140	1	TYR2_STRLN
36	7	1.3	149	1	ARGR_BACLI
37	7	1.3	156	1	ECR3_MOUSE
38	7	1.3	166	1	HES5_RAT
39	7	1.3	193	1	SAP3_MOUSE
40	7	1.3	205	1	EPA1_HUMAN
41	7	1.3	205	1	EPA1_MOUSE
42	7	1.3	205	1	EPA1_RAT
43	7	1.3	206	1	YMO8_YEAST
44	7	1.3	212	1	NOG2_BRARE
45	7	1.3	220	1	TRBP_RHISN

0921X0 mus musculus
P55048 streptomyce
O86130 bacillus 11
O35290 mus musculus
Q03062 rattus norv
Q60648 mus musculus
P20827 homo sapien
P52793 mus musculus
P97553 rattus norv
Q04019 saccharomyc
Q9W740 brachydanio
P55403 rhizobium s

RESULT 1
SPRT2_MOUSE
ID
AC Q9WU03; Q9WU04; Q9WU05;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR ACTIVATOR INHIBITOR TYPE 2) (HAI-2).
GN SPINT2 OR HAI2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=9ALB/C;
RX MEDLINE=99160423; PubMed=10049781;
RA Itoh H., Kataoka H., Hamauna R., Kitamura N., Koono M.;
RT "Hepatocyte growth factor activator inhibitor type 2 lacking the first Kunitz-type serine proteinase inhibitor domain is a predominant product in mouse but not in human."
RT Biochem. Biophys. Res. Commun. 255:740-748(1999).
CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: ISOFORM 2 IS MORE PREDOMINANTLY EXPRESSED THAN ISOFORM 1.
CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF099016; AAD22172.1; -
CC EMBL; AF099019; AAD22173.1; -
CC EMBL; AF099020; AAD22174.1; -
CC HSSP; P05067; ITAW.
CC MGD; MGI:1338031; Spint2.
CC InterPro: IPR002223; Kunitz_BPTI.
CC Pfam: PF00014; Kunitz_BPTI; 2.
CC PRINTS; PR00759; BASICPTASE.
CC SMART; SM00131; KU; 2.
CC PROSITE; PS00280; BPTI_KUNITZ_1; 2.
CC PROSITE; PS50279; BPTI_KUNITZ_2; 2.
CC Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane; Signal; Alternative splicing.
CC SIGNAL; 1 27
CC CHAIN 28 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.
CC DOMAIN 28 197 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 198 218 POTENTIAL.
 FT DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
 FT DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
 FT DISULFID 38 88 BY SIMILARITY.
 FT DISULFID 47 71 BY SIMILARITY.
 FT DISULFID 63 84 BY SIMILARITY.
 FT ACT_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 133 183 BY SIMILARITY.
 FT DISULFID 142 166 BY SIMILARITY.
 FT DISULFID 158 179 BY SIMILARITY.
 FT ACT_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 37 93 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT VASAPLIC 114 128 PROSADLSAELFN -> CFEVLSVALLEFLYA (IN ISOFORM 3).
 FT VASAPLIC 129 252 MISSING (IN ISOFORM 3).
 FT VASAPLIC 252 AA; 27914 MM; B2FF4B86924DAF8F CRC64;
 SQ SEQUENCE

Query Match
 Best Local Similarity 2.4%; Score 13; DB 1; Length 252;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 386 PFYVGGCEGNGNN 398
 DB 65 PFYVGGCEGNGNN 77

RESULT 2
 EFA2_MOUSE
 ID EFA2_MOUSE STANDARD; PRT; 209 AA.
 AC P52801;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE EPHRIN-A2 PRECURSOR (EPH-RELATED RECEPTOR TYROSINE KINASE LIGAND 6)
 DE (LEKR-6) (ELF-1) (CEK7-LIGAND) (CEK7-L).
 GN EFNA2 OR EPLG6 OR LERK6 OR ELFI OR EPL6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RA MEDLINE=95007776; PubMed=7522971;
 RA Cheng H.J., Flanagan J.G.;
 RT "Identification and cloning of ELF-1, a developmentally expressed
 RT ligand for the Mek4 and sek receptor tyrosine kinases.";
 RL Cell 79:157-168(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=95181289; PubMed=7876076;
 RA Shao H., Lou L., Pandey A., Verderame M.F., Slever D.A., Dixit V.M.;
 RT "cDNA cloning and characterization of a Cek7 receptor
 RT protein-tyrosine kinase ligand that is identical to the ligand
 RT (ELF-1) for the Mek-4 and sek receptor protein-tyrosine kinases.";
 RL J. Biol. Chem. 270:3467-3470(1995).
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
 CC EPHA5.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; U14941; AAA53636.1; -
 DR EMBL; U14752; AAA68520.1; -
 DR MGD; MGI:102707; efna2.
 DR InterPro; IPR001799; Ephrin.
 DR Pfam; PF00812; Ephrin; 1.
 DR ProDom; PD002533; Ephrin; 1.
 DR ProSITE; PS01299; EPHRIN; 1.
 DR Glycoprotein; GPI-anchor; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 209 EPHRIN-A2.
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 209 AA; 23586 MM; F1997545F2599ABC CRC64;

Query Match
 Best Local Similarity 1.8%; Score 10; DB 1; Length 209;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 RPLPLLL 14
 DB 6 RPLPLLL 15

RESULT 3
 EFA2_HUMAN
 ID EFA2_HUMAN STANDARD; PRT; 213 AA.
 AC O43921; 076020;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE EPHRIN-A2 PRECURSOR (EPH-RELATED RECEPTOR TYROSINE KINASE LIGAND 6)
 DE (LEKR-6) (HEK7-LIGAND) (HEK7-L).
 GN EFNA2 OR EPLG6 OR LERK6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98126446; PubMed=9465306;
 RA Carretti D.P., Nelson N.;
 RT "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3),
 RT mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (EPLG6):
 RT conservation of intron/exon structure.";
 RL Genomics 47:131-135(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Ganes J., Dangana L., Poundstone P.,
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
 RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,
 RA Carrino A.V.;
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=99045414; PubMed=9826538;
 RA Asheim H.C., Pedetour F., Grosgeorge J., Logtenberg T.;
 RT "Cloning, chromosomal mapping, and tissue expression of the gene
 RT encoding the human Eph-family kinase ligand ephrin-A2.";
 RL Biochem. Biophys. Res. Commun. 252:378-382(1998).
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
 CC EPHA5.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U92896; AAC39577.1; -.
DR EMBL; U92893; AAC39577.1; JOINED.
DR EMBL; U92894; AAC39577.1; JOINED.
DR EMBL; AC004258; AAC04896.1; -.
DR EMBL; AJ007292; CA07435.1; -.
DR MIM; 602756; -.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
KW Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 24
FT CHAIN 1 24
FT CARBOHYD 42 213
FT CARBOHYD 174 42
FT CARBOHYD 174 174
FT CARBOHYD 188 188
FT CONFLICT 6 6
FT CONFLICT 25 26
FT CONFLICT 29 30
FT CONFLICT 30 30
SQ SEQUENCE 213 AA; 23878 MW; 33C9FB1A8168B2D0 CRC64;

Query Match
Best Local Similarity 1.8%; Score 10; DB 1; Length 213;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 RPLPLLLLL 14
DB 6 RPLPLLLLL 15

RESULT 4
TKDL_BOVIN STANDARD; PRT; 164 AA.
ID ID TKDL_BOVIN STANDARD; PRT; 164 AA.
AC AC Q28201.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TROPHOBLAST KUNITZ DOMAIN PROTEIN 1 (TKDP-1) (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=941171738; PubMed=7510284;
RA Kriemer K.K., Duffy J.Y., Klemann S.W., Bixby J.A., Low B.G.,
RA Pope W.F., Roberts R.M.;
RA "Selective cloning of cDNA for secretory proteins of early embryos.
RT Identification of a transiently expressed kunitz domain protein from
RT preimplantation sheep trophoblast.";
RL J. Biol. Chem. 269:7255-7261(1994).
CC -1- FUNCTION: MAY PLAY A ROLE IN MEDIATING MATERNAL-CONCEPTUS
CC INTERACTIONS IN THE IMMEDIATE PREIMPLANTATION PERIOD. DOES NOT
CC SEEMS TO HAVE PROTEINASE INHIBITORY ACTIVITY (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION AT DAYS 17-19 OF
CC PREGNANCY.
CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L11344; AAA0795.1; -.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR SMART; SM00131; KU; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
KW Repeat; Glycoprotein.
FT DOMAIN 1 98
FT NON_TER 1 13
FT REPEAT 15 98
FT REPEAT 15 98
FT DOMAIN 107 157
FT DISULFID 107 157
FT DISULFID 116 140
FT DISULFID 132 153
FT ACT_SITE 117 118
FT ACT_SITE 78 78
FT CARBOHYD 78 78
SQ SEQUENCE 164 AA; 17884 MW; B95C2780C0B71717 CRC64;

Query Match
Best Local Similarity 1.6%; Score 9; DB 1; Length 164;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 391 GCEGNGNMF 399
DB 139 GCEGNGNMF 147

RESULT 5
TKDL_SHEEP STANDARD; PRT; 265 AA.
ID ID TKDL_SHEEP STANDARD; PRT; 265 AA.
AC AC Q29428.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TROPHOBLAST KUNITZ DOMAIN PROTEIN 1 PRECURSOR (TKDP-1).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=uterus, and Conceptus membrane;
RA MEDLINE=941171738; PubMed=7510284;
RA Kriemer K.K., Duffy J.Y., Klemann S.W., Bixby J.A., Low B.G.,
RA Pope W.F., Roberts R.M.;
RA "Selective cloning of cDNA for secretory proteins of early embryos.
RT Identification of a transiently expressed kunitz domain protein from
RT preimplantation sheep trophoblast.";
RL J. Biol. Chem. 269:7255-7261(1994).
CC -1- FUNCTION: MAY PLAY A ROLE IN MEDIATING MATERNAL-CONCEPTUS
CC INTERACTIONS IN THE IMMEDIATE PREIMPLANTATION PERIOD. DOES NOT
CC SEEMS TO HAVE PROTEINASE INHIBITORY ACTIVITY.
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN THE TROPHOCTODERM, WHICH
CC FORMS THE OUTER EPITHELIAL LAYER OF THE TROPHOBLAST.
CC -1- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION AT DAYS 14 AND 16 OF
CC PREGNANCY.
CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

DR EMBL; U00165; AAA19108.1; .
 DR EMBL; L11343; AAB46361.1; .
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR PRINTS: PR00759; BASICEPTASE.
 DR SMART: SM00131; KU; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
 KW Repeat; signal; Glycoprotein.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 265 TROPHOBLAST KUNITZ DOMAIN PROTEIN 1.
 FT DOMAIN 32 200 2 X 84 AA APPROXIMATE REPEATS.
 FT REPEAT 32 115 1.
 FT REPEAT 117 200 2.
 FT DOMAIN 208 258 BPTI/KUNITZ INHIBITOR.
 FT DISULFID 208 258 BY SIMILARITY.
 FT DISULFID 217 241 BY SIMILARITY.
 FT ACT SITE 233 254 BY SIMILARITY.
 FT ACT SITE 218 219 REACTIVE BOND (POTENTIAL).
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 265 AA; 28964 MW; D700805DF03E3132 CRC64;.

Query Match Best Local Similarity 1.6%; Score 9; DB 1; Length 265;
 Matches 9; Conservative 0; Pred. No. 0.57;
 Mismatches 0; Indels 0; Gaps 0;
 QY 391 GCEGNGNMF 399
 DB 240 GCEGNGNMF 248

RESULT 6
 ID IIP2_HUMAN STANDARD; PRT; 328 AA.
 AC P18065; Q14619;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 2 PRECURSOR (IGFBP-2)
 DE (IIP-2) (IGF-BINDING PROTEIN 2).
 GN IGFBP2 OR BP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=91293227; PubMed=171313;
 RA Agarwal N., Hsieh C.L., Sillis D., Swaroop M., Desai B., Francke U.,
 Swaroop A.;
 "Sequence analysis, expression and chromosomal localization of a
 gene, isolated from a subtracted human retina cDNA library, that
 encodes an insulin-like growth factor binding protein (IGFBP-2)."
 RT Exp. Eye Res. 52:549-561(1991).
 RN [2]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 40-77.
 RC MEDLINE=90368661; PubMed=1697583;
 RA Zapp J., Klefer M., Merryweather J., Maslarsz F., Bauer D., Born W.,
 Fischer J.A., Foresch E.R.;
 "Isolation from adult human serum of four insulin-like growth factor
 (IGF) binding proteins and molecular cloning of one of them that is
 increased by IGF I administration and in extrapancreatic tumor
 hypoglycemia."
 RT J. Biol. Chem. 265:14892-14898(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal liver;
 RX MEDLINE=90060007; PubMed=2479552;
 RA Binkert C., Landwehr J., Mary J.L., Schwander J., Heinrich G.;

RT "Cloning, sequence analysis and expression of a cDNA encoding a novel
 RT insulin-like growth factor binding protein (IGFBP-2)."
 RT EMBO J. 8:2497-2502(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=9124821; PubMed=1710112;
 RA Ehrenborg E., Vilhelmsdottir S., Bajajica S., Larsson C., Stern I.,
 Koch J., Brondum-Nielsen K., Luthman H.;
 "Structure and localization of the human insulin-like growth factor-
 RT binding protein 2 gene."
 RT Biochem. Biophys. Res. Commun. 176:1250-1255(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=92293159; PubMed=1376411;
 RA Binkert C., Margot J.B., Landwehr J., Heinrich G., Schwander J.;
 "Structure of the human insulin-like growth factor binding protein-2
 RT gene."
 RT Mol. Endocrinol. 6:826-836(1992).
 CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
 CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
 CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
 CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- MISCELLANEOUS: BINDS IGF-11 MORE THAN IGF-1.
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
 CC PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch)
 CC
 DR EMBL; S37730; AAB22308.1; .
 DR EMBL; S37712; AAB22308.1; JOINED.
 DR EMBL; S37722; AAB22308.1; JOINED.
 DR EMBL; S37726; AAB22308.1; JOINED.
 DR EMBL; M35410; AAA03246.1; .
 DR EMBL; X16302; CAA34373.1; .
 DR EMBL; M69241; AAA36048.1; .
 DR EMBL; M69237; AAA36048.1; JOINED.
 DR EMBL; M69239; AAA36048.1; JOINED.
 DR EMBL; M69240; AAA36048.1; JOINED.
 DR EMBL; A09809; CAA00862.1; .
 DR PIR; S05389; S05389.
 DR PIR; A35803; A35803.
 DR PIR; J01029; J01029.
 DR PIR; A41927; A41927.
 DR PIR; A53957; A53957.
 DR MIM; 146731; .
 DR InterPro: IPR000867; IGFBP.
 DR InterPro: IPR000716; Thyroglobulin_1.
 DR Pfam: PF00219; IGFBP; 1.
 DR Pfam: PF00086; thyroglobulin_1; 1.
 DR SMART: SM00121; IB; 1.
 DR SMART: SM00211; TY; 1.
 DR PROSITE: PS00222; IGF-BINDING; 1.
 DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
 KW Growth factor binding; signal.
 FT SIGNAL 1 39
 FT CHAIN 40 328
 FT DOMAIN 260 309 INSULIN-LIKE GROWTH FACTOR BINDING
 FT SITE 304 306 PROTEIN 2.
 FT CONFLICT 60 60 THYROGLOBULIN TYPE I.
 FT CONFLICT 320 320 CELL ATTACHMENT SITE.
 FT CONFLICT 323 323 P -> R (IN REF. 4).
 FT CONFLICT 323 323 R -> C (IN REF. 3).
 FT CONFLICT 323 323 H -> D (IN REF. 4).
 SQ SEQUENCE 328 AA; 35137 MW; 4E6BDFD805C8853 CRC64;

Query Match 1.68; Score 9; DB 1; Length 328;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PULPULLLL 14
 |||||||||
 DB 19 PULPULLLL 27

RESULT 7
 UPAR HUMAN STANDARD; PRT: 335 AA.
 AC 003405;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE UROKINASE PLASMINOGEN ACTIVATOR SURFACE RECEPTOR, GPI-ANCHORED FORM
 DE PRECURSOR (U-PAR) (MONOCYTE ACTIVATION ANTIGEN MO3) (CD87 ANTIGEN).
 CN PLAU OR UPAR OR MO3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-33.
 RX MEDLINE=90151622; PubMed=1689240;
 RA Roldan A.L., Cubellis M.V., Masucci M.T., Behrendt N., Lund L.R.,
 Danoe K., Appella E., Blas F.;
 RT "Cloning and expression of the receptor for human urokinase
 RT plasminogen activator, a central molecule in cell surface, plasmin
 RT dependent proteolysis.";
 RL EMBO J. 9:467-474(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92268511; PubMed=1316922;
 RA Min H.Y., Semnani R., Mizukami I.F., Watt K., Todd R.F. III, Liu D.Y.;
 RT "cDNA for MO3, a monocyte activation antigen, encodes the human
 RT receptor for urokinase plasminogen activator.";
 RL J. Immunol. 148:3636-3642(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94325536; PubMed=8049431;
 RA Casey J.R., Petranka J.G., Koltira J., Fleenor D.E., Rosse W.F.;
 RT "The structure of the urokinase-type plasminogen activator receptor
 RT gene.";
 RL Blood 84:1151-1156(1994).
 RN [4]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=90202929; PubMed=2156852;
 RA Behrendt N., Roenne E., Ploug M., Petri T., Loeber D., Nielsen L.S.,
 RA Schlenker W.-D., Blas F., Appella E., Danoe K.;
 RT "The human receptor for urokinase plasminogen activator. NH2-terminal
 RT amino acid sequence and glycosylation variants.";
 RL J. Biol. Chem. 265:6453-6460(1990).
 RN [5]
 RP SEQUENCE OF 106-116, AND CLEAVAGE BY U-PA.
 RX MEDLINE=97182578; PubMed=9030717;
 RA Hoeyer-Hansen G., Ploug M., Behrendt N., Roenne E., Danoe K.;
 RT "Cell-surface acceleration of urokinase-catalyzed receptor cleavage.";
 RL Eur. J. Biochem. 243:21-26(1997).
 RN [6]
 RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
 RX MEDLINE=93352546; PubMed=8394346;
 RA Ploug M., Kjalke M., Roenne E., Weidle U., Hoeyer-Hansen G., Danoe K.;
 RT "Localization of the disulfide bonds in the NH2-terminal domain of
 RT the cellular receptor for human urokinase-type plasminogen activator.
 RT A domain structure belonging to a novel superfamily of glycolipid-
 RT anchored membrane proteins.";
 RL J. Biol. Chem. 268:17539-17546(1993).
 CC -1- FUNCTION: ACTS AS A RECEPTOR FOR UROKINASE PLASMINOGEN ACTIVATOR.

CC PLAYS A ROLE IN LOCALIZING AND PROMOTING PLASMIN FORMATION.
 CC MEDIATES THE PROTEOLYSIS-INDEPENDENT SIGNAL TRANSDUCTION
 CC ACTIVATION EFFECTS OF U-PA. IT IS SUBJECT TO NEGATIVE-FEEDBACK
 CC REGULATION BY U-PA WHICH CLEAVES IT INTO AN INACTIVE FORM.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GPI-ANCHORED (SHOWN HERE) AND
 CC SECRETED; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 3 UPAR/LY6 DOMAINS.
 CC -1- DATABASE: NAME=PRO; NOTE=CD guide CD87 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd87.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X51675; CA35981.1; -;
 DR EMBL; M83246; AAA39862.1; -;
 DR EMBL; U09937; AAB60690.1; -;
 DR EMBL; U09931; AAB60690.1; JOINED.
 DR EMBL; U09933; AAB60690.1; JOINED.
 DR EMBL; U09933; AAB60690.1; JOINED.
 DR EMBL; U09935; AAB60690.1; JOINED.
 DR EMBL; U09936; AAB60690.1; JOINED.
 DR EMBL; A18757; CA01421.1; -;
 DR PIR; S12376; S12376.
 DR MIM; 173391; -;
 DR InterPro; IPR001526; LY6_UPAR.
 DR Pfam; PF00021; UPAR_LY6; 2.
 DR SMART; SM00134; LU; 3.
 DR PROSITE; PS00983; LY6_UPAR; 3.
 DR KMW Receptor; Signal; Glycoprotein; GPI-anchor; Repeat;
 KW Alternative splicing.
 FT SIGNAL 1 22
 FT CHAIN 23 7305
 FT FT
 FT PROPEP 7306 335
 FT DOMAIN 23 114
 FT DOMAIN 115 213
 FT DOMAIN 214 305
 FT CARBOHYD 74 74
 FT CARBOHYD 184 184
 FT CARBOHYD 194 194
 FT CARBOHYD 222 222
 FT CARBOHYD 225 225
 FT DISULFID 25 46
 FT DISULFID 28 34
 FT DISULFID 39 67
 FT DISULFID 93 98
 FT DISULFID 117 144
 FT DISULFID 120 127
 FT DISULFID 137 169
 FT DISULFID 175 192
 FT DISULFID 193 198
 FT DISULFID 216 244
 FT DISULFID 219 227
 FT DISULFID 237 263
 FT DISULFID 269 287
 FT DISULFID 288 293
 FT LIPID 305 305
 FT SITE 105 106
 FT SITE 111 112
 FT CONFLICT 28 28
 FT SEQUENCE 335 AA; 36978 MW; AB19633EA3DC77171 CRC64;

Query Match 1.68; Score 9; DB 1; Length 335;
 Best Local Similarity 100.0%; Pred. No. 0.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PLPLPLLL 14
 |||||
 DB 5 PLPLPLLL 13

RESULT 8
 PVR2_MOUSE STANDARD; PRT: 530 AA.
 ID PVR2_MOUSE P32507; 062096;
 AC 01-OCT-1993 (Rel. 27, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE POLIOVIRUS RECEPTOR RELATED PROTEIN 2 PRECURSOR (MURINE HERPESVIRUS ENTRY PROTEIN B) (MHEB) (NECTIN 2) (POLIOVIRUS RECEPTOR HOMOLOG).
 GN PVR2 OR PVS OR PVR OR MPH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=92219365; PubMed=1560525;
 RA Morrison M.E., Racanelli V.R.;
 RT "Molecular cloning and expression of a murine homolog of the human poliovirus receptor gene";
 RL J. Virol. 66:2807-2813(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM BETA).
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=94179228; PubMed=8132569;
 RA Aoki J., Koike S., Ise I., Sato-Yoshida Y., Nomoto A.;
 RT "Amino acid residues on human poliovirus receptor involved in interaction with poliovirus";
 RL J. Biol. Chem. 269:8431-8438(1994).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=99214397; PubMed=10196354;
 RA Shukla D., Rowe C.L., Dong Y., Racanelli V.R., Spear P.G.;
 RT "The murine homolog (Mph) of human herpesvirus entry protein B (HvB) mediates entry of pseudorabies virus but not herpes simplex virus types 1 and 2";
 RL J. Virol. 73:4493-4497(1999).
 CC -1- FUNCTION: RECEPTOR FOR ALPHAHERPESVIRUS (SUCH AS MURINE HSV) ENTRY INTO CELLS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA AND BETA (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: BRAIN, SPINAL CORD, SPLEEN, KIDNEY, HEART AND LIVER.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: M80206; AAA39734.1;
 DR EMBL: D26107; BAA05103.1;
 DR PIR: A38211; HLMSP3.
 DR MGD: MGI:97822; PVS.
 DR InterPro: IPR000581; ILVD_EBD.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR Pfam: PF00047; Ig_3.
 DR ProDom: PD002691; ILVD_EBD; 1.
 DR SMART: SM00406; IGV; 1.
 KW Immunoglobulin domain; Receptor; Transmembrane; Glycoprotein; Signal; Alternative splicing.

FT SIGNAL 1 31
 FT CHAIN 32 530
 FT DOMAIN 32 351
 FT TRANSMEM 352 372
 FT DOMAIN 373 530
 FT DOMAIN 47 138
 FT DOMAIN 167 236
 FT DOMAIN 267 327
 FT DISULFID 54 131
 FT DISULFID 174 229
 FT DISULFID 274 320
 FT CARBOHYD 128 138
 FT CARBOHYD 138 138
 FT CARBOHYD 315 315
 FT VARSPIC 339 467

POTENTIAL.
 POLIOVIRUS RECEPTOR RELATED PROTEIN 2.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 IG-LIKE V-TYPE DOMAIN.
 IG-LIKE C2-TYPE DOMAIN 1.
 IG-LIKE C2-TYPE DOMAIN 2.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).
 ESPSTAGAGTGGTGGTAAATATAVAGTGILICQQRKE
 ORLOADEEEELGPPSYKPPPKALPEEPMSQLTLLGA
 SEHSVPKTPYFAGVSCADQEMPRYHEPLTERSGPLTGA
 ATGILGP -> DTPQASRDVGPLVWGVAGTLLVLLAGFL
 ALLILGRRRRKSPGGSGDGRGSDPTQVYGGNGPVEM
 RSASPERMPRDPGDEEEDEEEMKAEKGLMPPHSPKDDM
 ESHLGSLSLRRAVY (IN ISOFORM ALPHA).
 MISSING (IN ISOFORM ALPHA).
 OED7IBPA2B231B6 CRC64;

Query Match 1.6%; Score 9; DB 1; Length 530;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PLPLPLLL 14
 |||||
 DB 17 PLPLPLLL 25

RESULT 9
 ITAS_MOUSE STANDARD; PRT: 1053 AA.
 ID ITAS_MOUSE P11688;
 AC 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE INTEGRIN ALPHA-5 PRECURSOR (FIBRONECTIN RECEPTOR ALPHA SUBUNIT) (INTEGRIN ALPHA-F) (VLA-5) (CD49E).
 GN ITGA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA Morrisey E., Dutt P., Patel V.;
 RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 645-1053 FROM N.A.
 RC STRAIN=BAH/C;
 RX MEDLINE=89235580; PubMed=2523953;
 RA Holers V.M., Ruff T.G., Parks D.L., McDonald J.A., Ballard L.L., Brown E.J.;
 RT "Molecular cloning of a murine fibronectin receptor and its expression during inflammation. Expression of VLA-5 is increased in activated peritoneal macrophages in a manner discordant from major histocompatibility complex class II";
 RL J. Exp. Med. 169:1589-1605(1989).
 CC -1- FUNCTION: INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRONECTIN AND FIBRINOGEN. IT RECOGNIZES THE SEQUENCE R-G-D IN ITS LIGANDS. WIDE HOMOLOGOUS FOR A NULL MUTATION OF THE ALPHA-5 SUBUNIT GENE DIE AT DAY E10-E11. THEY SHOW BOTH EXTRA-EMBRYONIC AND EMBRYONIC VASCULAR DEFECTS, AND SEVERE ABNORMALITIES IN THE DEVELOPMENT OF THE POSTERIOR TRUNK. MAY PLAY A ROLE IN THE SURVIVAL OF ADULT SKELETAL MUSCLE.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A

CC DISULFIDE BOND. ALPHA-5 ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X79003; CA55638.1; -;
 DR EMBL: X15203; CA33273.1; -;
 DR PIR: P10103; P10103.
 DR HSSP: P11215; 1A8X.
 DR MGD: MGI:96604; Itga5.
 DR InterPro: IPR000413; Integrin_alpha.
 DR Pfam: PF01839; FG-GAP; 5.
 DR Pfam: PF00357; Integrin_A; 1.
 DR PRINTS: PRO1185; INTEGRINA.
 DR SMART: SM00191; Int_alpha; 5.
 DR Prosite: PS00242; INTEGRIN_ALPHA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Extracellular matrix; Cytoskeleton; Calcium; Repeat.
 FT SIGNAL 1 44
 FT CHAIN 45 1053
 FT CHAIN 45 898
 FT CHAIN 899 1053
 FT DOMAIN 45 999
 FT TRANSMEM 1000 1025
 FT DOMAIN 1026 1053
 FT REPEAT 134 201
 FT REPEAT 202 259
 FT REPEAT 272 324
 FT REPEAT 326 390
 FT REPEAT 392 458
 FT REPEAT 456 508
 FT CA_BIND 337 345
 FT CA_BIND 404 412
 FT CA_BIND 468 476
 FT SITE 1024 1028
 FT DISULFID 102 111
 FT DISULFID 159 179
 FT DISULFID 195 208
 FT DISULFID 516 525
 FT DISULFID 531 587
 FT DISULFID 648 654
 FT DISULFID 721 734
 FT DISULFID 873 915
 FT DISULFID 920 925
 FT CARBOHYD 87 87
 FT CARBOHYD 185 185
 FT CARBOHYD 300 300
 FT CARBOHYD 310 310
 FT CARBOHYD 319 319
 FT CARBOHYD 527 527
 FT CARBOHYD 533 533
 FT CARBOHYD 596 596
 FT CARBOHYD 612 612
 FT CARBOHYD 678 678
 FT CARBOHYD 715 715
 FT CARBOHYD 727 727
 FT CARBOHYD 776 776
 FT CARBOHYD 872 872
 SO SEQUENCE 1053 AA: 115056 MW: DB872D11AC1755A6 CRC64;

Qy 6 PULPLLLL 14
 Db 27 PULPLLLL 35
 RESULT 10
 ID PULP_MOUSE STANDARD: PRT; 1238 AA.
 AC 064455;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE ETA PRECURSOR (EC 3.1.3.48) (R-PTP-ETA)
 DE (HPTP BETA-LIKE TYROSINE PHOSPHATASE).
 GN PTPR OR BTP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MRJ-LPR/LPR;
 RX MEDLINE=96140699; PubMed=8549806;
 RA Kuramochi S., Matsuda S., Matsuda Y., Saitoh T., Ohsugi M.,
 RA Yamamoto T.;
 RT "Molecular cloning and characterization of Byp, a murine
 RT receptor-type tyrosine phosphatase similar to human DEP-1";
 RL FEBS Lett. 378:7-14(1996).
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN EVERY TISSUE EXAMINED.
 CC -1- SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D45212; BA08146.1; -;
 DR HSSP: P18052; 1YFO.
 DR MGD: MGI:104574; Ptpn1.
 DR InterPro: IPR001777; FN_III.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_prot_phphatase.
 DR Pfam: PF00041; fn3; 6.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR PRINTS: PRO0700; PTPYPPHTASE.
 DR SMART: SM00060; FN3; 5.
 DR SMART: SM00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS00506; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS00505; TYR_PHOSPHATASE_ptp; 1.
 KW Signal; Glycoprotein; Transmembrane; Repeat; Hydrolase.
 FT SIGNAL 1 28
 FT CHAIN 29 1238
 FT DOMAIN 29 876
 FT TRANSMEM 877 897
 FT DOMAIN 898 1238
 FT DOMAIN 121 260
 FT DOMAIN 268 348
 FT DOMAIN 356 434
 FT DOMAIN 442 518
 FT DOMAIN 529 608
 FT DOMAIN 966 1238
 FT AC SITE 1140 1140
 FT CARBOHYD 62 62
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 POTENTIAL.
 PROTEIN-TYROSINE PHOSPHATASE ETA.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 FIBRONECTIN TYPE-III 1.
 FIBRONECTIN TYPE-III 2.
 FIBRONECTIN TYPE-III 3.
 FIBRONECTIN TYPE-III 4.
 FIBRONECTIN TYPE-III 5.
 FIBRONECTIN TYPE-III 6.
 PROTEIN-TYROSINE PHOSPHATASE.
 BY SIMILARITY.

Query Match 1.6%; Score 9; DB 1; Length 1053;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 379 379 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 538 538 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 572 572 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 662 662 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 668 668 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1238 AA; 136782 MW; 939479EDC8016835 CRC64;

```

Query Match 1.6%; Score 9; DB 1; Length 1238;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 7 LLLPLLLLR 15
DB 21 LLLPLLLLR 29

```

```

RESULT 11
CAL7_HUMAN STANDARD; PRT; 2944 AA.
ID 002358; Q14054; Q16507;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(VII) CHAIN PRECURSOR (LONG-CHAIN COLLAGEN) (LC
DE COLLAGEN).
GN COL7A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP MEDLINE=9437588; PubMed=8051117;
RX MEDLINE=9437588; PubMed=8051117;
RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen. Complete primary sequence of the
RT alpha 1(VII) chain and identification of intragenic polymorphisms.";
RN J. Biol. Chem. 269:20256-20262(1994).
RP SEQUENCE OF 128-1493 FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=93338437; PubMed=1307247;
RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,

```

```

RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E.,
RA Uitto J.;
RT "The large non-collagenous domain (NC-1) of type VII collagen is
RT amino-terminal and chimeric. Homology to cartilage matrix protein,
RT the type III domains of fibronectin and the A domains of von
RT Wilbrand factor";
RN Hum. Mol. Genet. 1:475-481(1992).
RP SEQUENCE OF 815-1439 FROM N.A.
RX MEDLINE=9334380; PubMed=1871109;
RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W.,
RA Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;
RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
RT gene";
RN Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
RP SEQUENCE OF 369-1255 FROM N.A.
RX MEDLINE=93107742; PubMed=1469284;
RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisyann P.S.,
RA Cook M.E., Wright J., Briggaman R.A., Hunt S.W. III;
RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain
RT extracellular matrix";
RN J. Invest. Dermatol. 99:691-696(1992).
RP SEQUENCE OF 340-675 FROM N.A.
RX TISSUE-Keratinocytes;
RA MEDLINE=92231902; PubMed=1567409;
RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
RT "Molecular cloning and characterization of type VII collagen cDNA";
RN Biochem. Biophys. Res. Commun. 183:958-963(1992).
RP SEQUENCE OF 2395-2944 FROM N.A.
RX MEDLINE=93271985; PubMed=8439916;
RA Greenspan D.S.;
RT "The carboxyl-terminal half of type VII collagen, including the non-
RT collagenous NC-2 domain and intron/exon organization of the
RT Hum. Mol. Genet. 2:273-278(1993).
RP SEQUENCE OF 1-87 FROM N.A.
RX TISSUE-Placenta;
RA MEDLINE=94375010; PubMed=8088784;
RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,
RA Uitto J., Greenspan D.S.;
RT "Structural organization of the human type VII collagen gene (COL7A1),
RT composed of more exons than any previously characterized gene";
RN Genomics 21:169-179(1994).
RP REVIEW ON DEB VARIANTS.
RX MEDLINE=98041696; PubMed=9375848;
RA Jaervikallio A., Pulkkinen L., Uitto J.;
RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in
RT the type VII collagen gene (COL7A1)";
RN Hum. Mutat. 10:338-347(1997).
RP VARIANT DEB LYS-2798.
RX MEDLINE=93291877; PubMed=8513326;
RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y.,
RA Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;
RT "A missense mutation in type VII collagen in two affected siblings
RT with recessive dystrophic epidermolysis bullosa";
RN Nat. Genet. 4:62-66(1993).
RP VARIANT DEB SER-2040.
RX MEDLINE=94224777; PubMed=8170945;
RA Christiano A.M., Ryyanen M., Uitto J.;
RT "Dominant dystrophic epidermolysis bullosa: identification of a
RT gly-->ser substitution in the triple-helical domain of type VII
RT collagen";
RN Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).
RP VARIANT PEB-DEB CYS-2623.

```

RX MEDLINE-96081220; PubMed-8541842;
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;
 RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and
 RT identification of a glycine-to-cysteine substitution in the triple-
 RT helical domain of type VII collagen.";
 RL Hum. Mol. Genet. 4:1579-1583(1995).
 RN [12]
 RP VARIANT DDB ARG-2043.
 RX MEDLINE-95164985; PubMed-7861014;
 RA Christiano A.M., Morriconi A., Paradisi M., Angelo C., Mazzanti C.,
 RA Cavallieri R., Uitto J.;
 RT "A glycine-to-arginine substitution in the triple-helical domain of
 RT type VII collagen in a family with dominant dystrophic epidermolysis
 RT bullosa.";
 RL J. Invest. Dermatol. 104:438-440(1995).
 RN [13]
 RP VARIANTS RDEB AND DDEB.
 RX MEDLINE-96620218; PubMed-8644729;
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
 RT "Glycine substitutions in the triple-helical region of type VII
 RT collagen result in a spectrum of dystrophic epidermolysis bullosa
 RT phenotypes and patterns of inheritance.";
 RL Am. J. Hum. Genet. 58:671-681(1996).
 RN [14]
 RP VARIANT RDEB ARG-2575.
 RX MEDLINE-96154068; PubMed-8592061;
 RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
 RT genotype/phenotype correlation in a case of moderate clinical
 RT severity.";
 RL J. Invest. Dermatol. 106:119-124(1996).
 RN [15]
 RP VARIANT RDEB ARG-1782.
 RX MEDLINE-96183562; PubMed-8618018;
 RA Christiano A.M., McGrath J.A., Uitto J.;
 RT "Influence of the second COL7A1 mutation in determining the
 RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 106:766-770(1996).
 RN [16]
 RP VARIANT RDEB ASP-2073.
 RX MEDLINE-96310789; PubMed-8757758;
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
 RA Uitto J., Pope F.M., Eady R.A.J.;
 RT "Clinicopathological correlations of compound heterozygous COL7A1
 RT mutations in recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 107:171-177(1996).
 RN [17]
 RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 & R-2575.
 RX MEDLINE-97465605; PubMed-9326325;
 RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
 RA Freltag S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,
 RA de Prost Y.;
 RT "Characterization of 18 new mutations in COL7A1 in recessive
 RT dystrophic epidermolysis bullosa provides evidence for distinct
 RT molecular mechanisms underlying defective anchoring fibril
 RT formation.";
 RL Am. J. Hum. Genet. 61:599-610(1997).
 RN [18]
 RP VARIANT RDEB ARG-1652.
 RX MEDLINE-96106792; PubMed-9444387;
 RA Csechnamt-Friedman P.B., Karpaci S., Horvath A., Christiano A.M.;
 RT "Identification of a glycine substitution and a splice site mutation
 RT in the type VII collagen gene in a proband with milds recessive
 RT dystrophic epidermolysis bullosa.";
 RL Arch. Dermatol. Res. 289:640-645(1997).
 RN [19]
 RP VARIANTS DEB ARG-2009 AND ARG-2043.
 RX MEDLINE-97358588; PubMed-9215684;
 RA Winberg J.-O., Hammami-Hausasi N., Nilssen O., Anton-Lamprecht I.,
 RA Naylor S.L., Keirachner K., Zimmermann M., Kratje P.,
 RA Gedde-Dahl T., Jr., Bruckner-Tuderman L.;
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
 RT a splice site mutation in combination with a missense mutation in the

RT COL7A1 gene.";
 RL Hum. Mol. Genet. 6:1125-1135(1997).
 RN [20]
 RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
 RX MEDLINE-96834662; PubMed-9668111;
 RA Hammami-Hausasi N., Schumann H., Ragnunath M., Kilgus O., Luethi U.,
 RA Luger T., Bruckner-Tuderman L.;
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in
 RT intracellular accumulation of collagen VII, loss of anchoring
 RT fibrils, and skin blistering.";
 RL J. Biol. Chem. 273:19228-19234(1998).
 RN [21]
 RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.
 RX MEDLINE-98410969; PubMed-9740253;
 RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Uitto J.;
 RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
 RT bullosa.";
 RL J. Invest. Dermatol. 111:534-537(1998).
 RN [22]
 RP VARIANT RDEB ARG-1347.
 RX MEDLINE-99019477; PubMed-9804332;
 RA Terraccina M., Posteraro P., Schubert M., Sonego G., Atzori F.,
 RA Zamburano G., Bruckner-Tuderman L., Castiglia D.;
 RT "Compound heterozygosity for a recessive glycine substitution and a
 RT splice site mutation in the COL7A1 gene causes an unusually mild form
 RT of localized recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 111:744-750(1998).
 RN [23]
 RP VARIANTS DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.
 RX MEDLINE-99072663; PubMed-9856843;
 RA Rouan F., Pulkkinen L., Jonkman M.F., Bauer J.W.,
 RA

Query Match 1.6%; Score 9; DB 1; Length 2944;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 384 CHPFYGGC 392
 |||||||||
 DB 2904 CHPFYGGC 2912

RESULT 12
 ID ACH_MOUSE
 AC ACH_MOUSE STANDARD; PRT: 27 AA.
 AC 070174;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA-4 CHAIN PRECURSOR
 DE (FRAGMENT).
 GN CHRNA4 OR ACRA4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98420182; PubMed-9749753;
 RA Watanabe H., Zoli M., Changeux J.-P.;
 RT "Promoter analysis of the neuronal nicotinic acetylcholine receptor
 RT alpha 4 gene: methylation and expression of the transgene.";
 RL Eur. J. Neurosci. 10:2244-2253(1998).
 CC -I- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
 CC EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
 CC LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
 CC MEMBRANE.
 CC -I- SUBUNIT: NEURONAL ACHR SEEMS TO BE COMPOSED OF TWO DIFFERENT
 CC TYPE OF SUBUNITS, ALPHA AND NON-ALPHA (BETA).
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -I- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AB010002; BAA25752.1; -.
DR MGD: MGI:87888; Chn24.
DR InterPro: IPR001175; Neur_channel.
DR PROSITE: PS00236; NEUROTR_ION_CHANNEL; PARTIAL.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
KW Transmembrane; Multigene family.
FT SIGNAL 1 >27 POTENTIAL.
FT NON_TER 27
SQ SEQUENCE 27 AA; 2680 MW; FAE7C1C9664F703A CRC64;

Query Match 1.5%; Score 8; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LRLPLLLL 14
Db 14 LRLPLLLL 21

RESULT 13
APC2_CAVPO STANDARD; PRT; 100 AA.

AC P27916;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE APOLIPOPROTEIN C-II PRECURSOR (APO-CII).
GN APOC2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathii; Cavillidae; Cavia.
OX NCBI_Taxid=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91154195; PubMed=1999402;
RA Andersson Y., Thelander L., Bengtsson-Olivera G.;
RT "Demonstration of apolipoprotein CII in guinea pigs. Functional
RT characteristics, cDNA sequence, and tissue expression."
RL J. Biol. Chem. 266:4074-4080(1991).
CC -I- FUNCTION: APOC-II IS A COMPONENT OF THE VERY LOW DENSITY
CC LIPOPROTEIN (VLDL) FRACTION IN PLASMA, AND IS AN ACTIVATOR OF
CC SEVERAL TRIACYLGLYCEROL LIPASES. THE ASSOCIATION OF APOC-II WITH
CC PLASMA CHYLOMICRONS, VLDL, AND HDL IS REVERSIBLE. A FUNCTION OF
CC THE SECRETION AND CATABOLISM OF TRIGLYCERIDE-RICH LIPOPROTEINS,
CC AND CHANGES RAPIDLY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M59913; AAA37031.1; -.
DR PIR: A38685; A38685.
KW Chylomicron; VLDL; Plasma; Lipid transport; Lipid degradation; Signal.
FT SIGNAL 1 25
FT CHAIN 26 100
SQ SEQUENCE 100 AA; 10984 MW; BBCF8DB52FC9E9BD CRC64;

Query Match 1.5%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LRLPLLLL 14
Db 12 LRLPLLLL 19

RESULT 14
YA84_AQUAE STANDARD; PRT; 122 AA.
AC 067176;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_1084.
GN AQ_1084.

OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_Taxid=65363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;

RX MEDLINE=9819666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000722; AAC07139.1; -.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 122 AA; 14488 MW; 5EB2F61D8554F276 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 530 KILLELLEK 537
Db 87 KILLELLEK 94

RESULT 15
RMP3_HUMAN STANDARD; PRT; 148 AA.

AC 060836;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RECEPTOR ACTIVITY-MODIFYING PROTEIN 3 PRECURSOR (CRMR ACTIVITY-
DE MODIFYING-PROTEIN 3) (CALCITONIN-RECEPTOR-LIKE RECEPTOR ACTIVITY-
DE MODIFYING-PROTEIN 3).
GN RAMP3.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.

KC TISSUE=Brain;
RX MEDLINE=98282119; PubMed=9620797;
RA McLatchie L.M., Fraser N.J., Main M.J., Wise A., Brown J.,
RA Thompson N., Solari R., Lee M.G., Foord S.M.,
RT "RAMPs regulate the transport and ligand specificity of the
RT calcitonin-receptor-like receptor."

```

RL  Neture 393:333-339(1998).
RN  [2]
RP  SEQUENCE OF 21-148 FROM N.A.
RA  Walker C., Bauer C., Smith R.
RL  Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: REQUIRED TO TRANSPORT CALCITONIN-RECEPTOR-LIKE RECEPTOR
CC  (CRLR) TO THE PLASMA MEMBRANE.
CC  -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC  -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN LUNG, BREAST, IMMUNE
CC  SYSTEM AND FETAL TISSUES.
CC  -1- SIMILARITY: BELONGS TO THE RAMP FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AJ001016; CAA04474.1; -.
DR  EMBL; AC004844; AAD0854.1; -.
DR  MIM; 605155; -.
KW  Signal; Transmembrane; Transport; Receptor.
FT  SIGNAL      1      23      POTENTIAL.
FT  CHAIN       24     148     RECEPTOR ACTIVITY-MODIFYING PROTEIN 3.
FT  DOMAIN      24     118     EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM    119     138     POTENTIAL.
FT  DOMAIN      139     148     CYTOPLASMIC (POTENTIAL).
SQ  SEQUENCE    148 AA; 16518 MW; EEE312496EF513C CRC64;

Query Match      1.5%; Score 8; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy  7 LLPLLLLL 14
    |          |
Db  11 LLPLLLLL 18

```

Search completed: February 26, 2002, 01:23:06
 Job time: 311 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:16:15 : Search time 21.58 Seconds
(without alignments)
1934.368 Million cell updates/sec

Title: US-09-819-136-2

Perfect score: 548
Sequence: 1 MPALRPPLPLLLRLRSGA.....KRIELLEKQACELNRFQD 548

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	13	2.4	252	2 JG0185	hepatocyte growth
2	10	1.8	209	2 A54984	ELF-1 protein prec
3	9	1.6	18	2 I52614	u-plasminogen acti
4	9	1.6	213	2 JEO322	ephrin-A2 - human
5	9	1.6	265	2 A53390	Kunitz-type protei
6	9	1.6	281	2 S39495	u-plasminogen acti
7	9	1.6	328	1 A41927	Insulin-like growt
8	9	1.6	335	2 A39743	u-plasminogen acti
9	9	1.6	422	2 S32357	glial growth facto
10	9	1.6	467	1 HUKSP3	poliovirus recepto
11	9	1.6	530	2 A54237	integrin alpha-5 c
12	9	1.6	1053	2 S44250	poliovirus recepto
13	9	1.6	1238	2 S68700	type VII collagen
14	9	1.6	1549	2 I48103	Hrnp beta-like tyr
15	9	1.6	2055	2 T00093	hypothetical prote
16	9	1.6	2944	2 A54849	collagen alpha 1(V
17	8	1.5	100	2 A38685	apolipoprotein C-I
18	8	1.5	106	2 T06479	proline/leucine-ri
19	8	1.5	142	2 E70393	hypothetical prote
20	8	1.5	147	2 JC7237	receptor-activit
21	8	1.5	147	2 JC7263	receptor activity-
22	8	1.5	184	2 S10125	alpha-2u-globulin
23	8	1.5	249	2 T35589	probable secreted
24	8	1.5	264	2 B82619	exodeoxyribonuclea
25	8	1.5	281	2 D72351	hypothetical prote
26	8	1.5	354	2 G75619	probable hemin ABC
27	8	1.5	507	2 T31179	hypothetical prote
28	8	1.5	527	2 A32469	80K protein H prec
29	8	1.5	546	2 G72764	hypothetical prote

30	8	1.5	625	2 A26456	nicotinic acetylch
31	8	1.5	662	2 A45155	mucin FIM-C.1 - Af
32	8	1.5	672	2 B84782	probable receptor-
33	8	1.5	699	2 JC6132	protein-tyrosine-p
34	8	1.5	700	1 S12053	protein-tyrosine-p
35	8	1.5	744	2 A45622	surface antigen gp
36	8	1.5	829	2 B86467	hypothetical prote
37	8	1.5	936	2 B64567	cytochrome c b1oge
38	8	1.5	936	2 H71862	probable cytochrom
39	8	1.5	1062	2 T46444	hypothetical prote
40	8	1.5	1321	2 T00382	hypothetical prote
41	8	1.5	6805	2 S20901	titin - rabbit (fr
42	8	1.5	26926	1 I38344	titin, cardiac mus
43	7	1.3	24	2 I67553	monocyte chemotact
44	7	1.3	45	2 C71360	hypothetical prote
45	7	1.3	51	2 S64676	acetylcholinestera

ALIGNMENTS

RESULT 1
JG0185
hepatocyte growth factor activator inhibitor type 2 - mouse
C.Species: Mus musculus (house mouse)
C.Date: 23-Jul-1999 #sequence=revision 23-Jul-1999 #ext_change 11-May-2000
C.Accession: JG0185
R.Itoh, H.; Kataoka, H.; Hamasuna, R.; Kitamura, N.; Kono, M.
Biochem. Biophys. Res. Commun. 255, 740-748, 1999
A.Title: Hepatocyte growth factor activator inhibitor type 2 lacking the first kunitz
A.Reference number: JG0185; MUID:99160423
A.Accession: JG0185
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-252 <IT0>
A.Cross-references: GB:AF099016
C.Superfamily: animal kunitz-type proteinase inhibitor homology
F.I33-183/Domain: animal kunitz-type proteinase inhibitor homology <BP1>

Query Match 2.4%; Score 13; DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 386 PEVYGGCEGNGNN 398
DB 65 PEVYGGCEGNGNN 77
RESULT 2
A54984
ELF-1 protein precursor - mouse
N.Alternate names: Cerk7 ligand
C.Species: Mus musculus (house mouse)
C.Date: 11-Nov-1994 #sequence=revision 11-Nov-1994 #ext_change 29-Sep-1999
C.Accession: A54984; A55873
R.Cheng, H.J.; Flanagan, J.G.
Cell 79, 157-168, 1994
A.Title: Identification and cloning of ELF-1, a developmentally expressed ligand for
A.Reference number: A54984; MUID:95007776
A.Accession: A54984
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-209 <CHE>
A.Cross-references: GB:U14941; NID:G558836; PIDN:AA53636.1; PID:G558837
R.Shao, H.; Lou, L.; Pandey, A.; Verderame, M.F.; Stever, D.A.; Dlxlt, V.M.
J. Biol. Chem. 270, 3467-3470, 1995
A.Title: CDNA cloning and characterization of a Cerk7 receptor protein-tyrosine kinase
A.Accession: A55873; MUID:95181289
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-209 <SHA>

A:Cross-references: GB:U14752; NID:668186; PIDN:AAA6520.1; PID:6681887
 C:Superfamily: axon guidance signal protein
 C:Keywords: lipoprotein; membrane protein

Query Match 1.8%; Score 10; DB 2; Length 209;
 Best Local Similarity 100.0%; Pred. No. 0.099;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 PLPLPLLLL 14
 Db 6 PLPLPLLLL 15

RESULT 3
 152614
 u-plasminogen activator receptor precursor - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 21-Jul-2000
 C:Accession: 152614
 R:Sortavia, E.; Grebe, A.; De Luca, P.; Hellin, K.; Suh, T.T.; Degen, J.L.; Blas, F.
 Blood 86, 624-635, 1995
 A:Title: A conserved TATA-less proximal promoter drives basal transcription from the uro
 A:Reference number: 152614; MUID:95329719
 A:Accession: 152614
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-18 <RES>
 A:Cross-references: GB:S78532; NID:9999307; PIDN:AAU14289.1; PID:94261989
 C:Genetics:
 A:Gene: UPAR
 C:Superfamily: urokinase-type plasminogen activator receptor; Ly-6 homology

Query Match 1.6%; Score 9; DB 2; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 PLPLPLLLL 14
 Db 5 PLPLPLLLL 13

RESULT 4
 JE0322
 ephrin-A2 - human
 C:Species: Homo sapiens (man)
 C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
 C:Accession: JE0322
 R:Aashheim, H.; Pedoutour, F.; Grosgeorge, J.; Logtenberg, T.
 Biochem. Biophys. Res. Commun. 252, 378-382, 1998
 A:Title: Cloning, chromosomal mapping, and tissue expression of the gene encoding the huma
 A:Reference number: JE0322; MUID:99045414
 A:Accession: JE0322
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-213 <AAS>
 A:Cross-references: GB:AJ007297; NID:93688367; PIDN:CAA07435.1; PID:93688368
 C:Superfamily: axon guidance signal protein

Query Match 1.6%; Score 9; DB 2; Length 213;
 Best Local Similarity 100.0%; Pred. No. 0.95;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 PLPLPLLLL 14
 Db 7 PLPLPLLLL 15

RESULT 5
 A53390
 Kunitz-type proteinase inhibitor precursor - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 05-Nov-1999
 C:Accession: A53390
 R:Kramer, K.K.; Duffly, J.Y.; Klemann, S.W.; Bixby, J.A.; Low, B.G.; Pope, W.F.; Rober
 J. Biol. Chem. 269, 7255-7261, 1994
 A:Title: Selective cloning of cDNA for secretory proteins of early embryos. Identific
 A:Reference number: A53390; MUID:94171738
 A:Accession: A53390
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-265 <KRA>
 A:Cross-references: GB:U00165; NID:9501024; PIDN:AA19108.1; PID:9392040
 C:Superfamily: animal Kunitz-type proteinase inhibitor homology
 C:Keywords: serine proteinase inhibitor
 F:208-258/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Query Match 1.6%; Score 9; DB 2; Length 265;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 391 GCEGNGNMF 399
 Db 240 GCEGNGNMF 248

RESULT 6
 S39495
 u-plasminogen activator receptor form 2 precursor - human
 N:Alternate names: urokinase-type plasminogen activator receptor form 2
 C:Species: Homo sapiens (man)
 C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 23-Jul-1999
 C:Accession: S39495; S34462
 R:Pyke, C.; Eriksen, J.; Solberg, H.; Schnack Nielsen, B.; Kristensen, P.; Lund, L.R.
 submitted to the EMBL Data Library, December 1993
 A:Description: An alternatively spliced variant of mRNA for the human receptor for ur
 A:Reference number: S39495
 A:Accession: S39495
 A:Molecule type: mRNA
 A:Residues: 1-281 <PKY1>
 A:Cross-references: EMBL:X74039; NID:9456192; PIDN:CAA52191.1; PID:9433901
 R:Pyke, C.; Eriksen, J.; Solberg, H.; Schnack Nielsen, B.; Kristensen, P.; Lund, L.R.
 FEBS Lett. 326, 69-74, 1993
 A:Title: An alternatively spliced variant of mRNA for the human receptor for urokinas
 A:Reference number: S34461; MUID:93314820
 A:Accession: S34462
 A:Molecule type: mRNA
 A:Residues: 246-281 <PKY2>
 A:Cross-references: GB:U09347; GB:S63673; NID:9483827; PIDN:AA17978.1; PID:9483828
 C:Superfamily: urokinase-type plasminogen activator receptor; Ly-6 homology
 C:Keywords: alternative splicing; receptor
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-281/Product: u-plasminogen activator receptor #status predicted <MAT>
 F:115-206/Domain: Ly-6 homology <LY6>

Query Match 1.6%; Score 9; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 PLPLPLLLL 14
 Db 5 PLPLPLLLL 13

RESULT 7
 A41927
 insulin-like growth factor-binding protein 2 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1993 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
 C:Accession: A41927; U01029; S05389; A35803; C54651
 R:Binkert, C.; Margot, J.B.; Landwehr, J.; Heinrich, G.; Schwander, J.
 Mol. Endocrinol. 6, 826-836, 1992

```

A:Title: Structure of the human insulin-like growth factor binding protein-2 gene.
A:Reference number: A41927; MUID:92233159
A:Accession: A41927
A:Molecule type: DNA
A:Residues: 1-328 <B1N>
A:Cross-references: GB:537730; NID:g250198; PIDN:AAB23308.1; PID:g250199
A:Experimental source: Placenta
A:Note: sequence extracted from NCBI backbone (NCBIN:106610, NCBIN:106613, NCBIN:106615,
R:Enrenborg, E.; Vilhelmsdottir, S.; Bajlicca, S.; Larsson, C.; Stern, I.; Koch, J.; Brö
Blochem. Biophys. Res. Commun. 176, 1250-1255, 1991
A:Title: Structure and localization of the human insulin-like growth factor-binding prot
A:Reference number: JQ1029; MUID:91248211
A:Accession: JQ1029
A:Molecule type: DNA
A:Residues: 1-59, 'R', 61-322, 'D', 324-328 <ENR>
A:Cross-references: GB:M69237
A:Note: the authors translated the codon CGC for residue 60 as Pro and the codon GAC for
R:Agarwal, N.; Hsieh, C.L.; Sills, D.; Swaroop, M.; Desai, B.; Francke, U.; Swaroop, A.
Exp. Eye Res. 52, 549-561, 1991
A:Title: Sequence analysis, expression and chromosomal localization of a gene, isolated
).
A:Reference number: A53957; MUID:91293227
A:Accession: A53957
A:Molecule type: mRNA
A:Residues: 1-328 <AGA>
A:Cross-references: GB:M35410; NID:g179476; PIDN:AAA03246.1; PID:g179477
A:Experimental source: retina
R:Blunkert, C.; Landwehr, J.; Mary, J.L.; Schwander, J.; Helmrich, G.
EMBO J. 8, 2497-2502, 1989
A:Title: Cloning, sequence analysis and expression of a cDNA encoding a novel insulin-l
A:Reference number: S05389; MUID:90060007
A:Accession: S05389
A:Molecule type: mRNA
A:Residues: 1-319, 'C', 321-328 <B12>
A:Cross-references: EMBL:X16302; NID:g33009; PIDN:CAA34373.1; PID:g33010
R:Zappf, J.; Klefer, M.; Merryweather, J.; Maslarsz, F.; Bauer, D.; Born, W.; Fischer, J.
J. Biol. Chem. 265, 14892-14899, 1990
A:Title: Isolation from adult human serum of four insulin-like growth factor (IGF) bindi
IG tumor hypoglycemia.
A:Reference number: A35803; MUID:90368661
A:Accession: A35803
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 119, 'P', 21-328 <ZAP>
R:Rognan, M.; Segovia, B.; Whitechurch, O.; Blinoux, M.
Growth Regul. 1, 125-130, 1991
A:Title: Purification from human cerebrospinal fluid of insulin-like growth factor bindi
A:Reference number: A54651; MUID:93091816
A:Accession: C54651
A:Molecule type: protein
A:Residues: 43-44, 'X', 46-47, 'X', 49-55, 'X', 57-61 <ROG>
A:Experimental source: cerebrospinal fluid
A:Note: sequence extracted from NCBI backbone (NCBIP:121072)
C:Genetics:
A:Gene: GDB:IGFBP2; IBP2
A:Cross-references: GDB:126735; OMIM:146731
A:Map position: 2q33-2q34
A:Intons: 151/1; 227/3; 274/3
C:Superfamily: Insulin-like growth factor binding protein 1; thyroglobulin type I repeat
C:Keywords: plasma
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-38/Product: insulin-like growth factor-binding protein 2 #status predicted <MAT>
F:230-309/Domain: thyroglobulin type I repeat homology <HY1>
F:304-306/Region: cell attachment (R-G-D) motif

Query Match 1.6%; Score 9; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 PLPLPLLL 14
| | | | | | | |
Db 19 PLPLPLLL 27

```

```

RESULT 8
A39743
U-plasminogen activator receptor form 1 precursor - human
N:Alternate names: monocyte activation antigen Mo3; urokinase-type plasminogen activa
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Jul-1999
C:Accession: S12376; A39743; A55242; S34461; I56138; S67529; I39206
R:Roldan, A.L.; Cuddebelli, M.V.; Masucci, M.T.; Behrendt, N.; Lund, L.R.; Dano, K.; Ap
EMBO J. 9, 467-474, 1990
A:Title: Cloning and expression of the receptor for human urokinase plasminogen activ
A:Reference number: S12376; MUID:90151622
A:Accession: S12376
A:Molecule type: mRNA
A:Residues: 1-335 <RDL>
A:Cross-references: GB:X51675; NID:g37604; PIDN:CAA35981.1; PID:g37605
R:Behrendt, N.; Ploug, M.; Patchy, L.; Houen, G.; Blas, F.; Dano, K.
J. Biol. Chem. 265, 7842-7847, 1991
A:Title: The ligand-binding domain of the cell surface receptor for urokinase-type pl
A:Reference number: A39743; MUID:91210310
A:Accession: A39743
A:Molecule type: protein
A:Residues: 23-24, 'X', 26-27, 'XX', 30-33, 'X', 35-38, 'X', 40, 'X', 42, 110-116, 'XX', 119, 'X' <
R:Behrendt, N.; Rønne, E.; Ploug, M.; Petri, T.; Lober, D.; Nielsen, L.S.; Schleuning
J. Biol. Chem. 265, 6453-6460, 1990
A:Title: The human receptor for urokinase plasminogen activator. NH-2-terminal amino
A:Reference number: A35242; MUID:90202929
A:Accession: A35242
A:Molecule type: protein
A:Residues: 23, 'X', 25-38, 'H', 40-43, 'X', 45, 'X', 47-51, 'X' <BE2>
A:Note: the authors show a tentative identification of His for residue 39
R:Pyke, C.; Eriksen, J.; Solberg, H.; Schnack Nielsen, B.; Kristensen, P.; Lund, L.R.
FEBS Lett. 326, 69-74, 1993
A:Title: An alternatively spliced variant of mRNA for the human receptor for urokinas
A:Reference number: S34461; MUID:93314820
A:Accession: S34461
A:Molecule type: mRNA
A:Residues: 246-335 <PYK>
R:Ploug, M.; Rønne, E.; Behrendt, N.; Jensen, A.L.; Blas, F.; Dano, K.
J. Biol. Chem. 266, 1926-1933, 1991
A:Title: Cellular receptor for urokinase plasminogen activator. Carboxyl-terminal pro
A:Reference number: A38677; MUID:91107700
A:Accession: A38677
A:Contents: annotation
R:Mih, H.Y.; Seman, R.; Mizukami, I.F.; Watt, K.; Todd, R.F.
J. Immunol. 148, 3636-3642, 1992
A:Title: cDNA for Mo3, a monocyte activation antigen, encodes the human receptor for
A:Reference number: I56138; MUID:92265511
A:Accession: I56138
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-335 <RES>
A:Cross-references: GB:M83246; NID:g188623; PIDN:AAA59862.1; PID:g188624
R:Mang, Y.; Dang, J.; Johnson, L.R.; Selhammer, J.J.; Doe, W.F.
Eur. J. Biochem. 227, 116-122, 1995
A:Title: Structure of the human urokinase receptor gene and its similarity to CD59 an
A:Reference number: S67529; MUID:95154279
A:Accession: S67529
A:Molecule type: DNA
A:Residues: 1-335 <MAN>
A:Note: the authors translated the codon CGC for residue 51 as Ala
R:Casey, J.R.; Petrankska, J.G.; Kottra, J.; Fleenor, D.E.; Rosse, W.F.
Blood 84, 1151-1156, 1994
A:Title: The structure of the urokinase-type plasminogen activator receptor gene.
A:Reference number: I39206; MUID:94325556
A:Accession: I39206
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-335 <RE2>
A:Cross-references: EMBL:U09937; NID:g533374; PIDN:AAB60690.1; PID:g533376
C:Genetics:
A:Gene: UPAR
A:Intons: 19/1; 56/1; 104/1; 158/1; 203/1; 252/1

```

C:Superfamily: urokinase-type plasminogen activator receptor; Ly-6 homology
 C:Keywords: alternative splicing; glycoprotein; membrane protein; phosphatidylinositol
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-304/Product: (or 23-305 or 23-306) u-plasminogen activator receptor #status experim
 F:23-107/Domain: Ly-6 homology <LY6A>
 F:115-206/Domain: Ly-6 homology <LY6B>
 F:214-301/Domain: Ly-6 homology <LY6C>
 F:74,184,194,222,255/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 9; DB 2; Length 335;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PLPLPLLLL 14
 |||||
 Db 5 PLPLPLLLL 13

RESULT 9
 S32357
 glial growth factor - human
 C:Species: Homo sapiens (man)
 C>Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 11-Jan-2000
 C:Accession: S32357
 R:Marchionni, M.A.; Goodheart, A.D.J.; Chen, M.S.; Berningham-McDonogh, O.; Kirk, C.; Her
 les, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; Mcburney, R.N.; Waterfield, M.
 Nature 362, 312-318, 1993
 A>Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the
 A:Reference number: S32357; MUID:93205115
 A:Accession: S32357
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-422 <MAR>
 A:Cross-references: GB:112260; NID:q292047; PIDN:AA59622.1; PID:q292048
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:363-402/Domain: EGF homology <EGF>

Query Match 1.6%; Score 9; DB 2; Length 422;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PLPLPLLLL 14
 |||||
 Db 31 PLPLPLLLL 39

RESULT 10
 H1MSP3
 poliovirus receptor homolog precursor - mouse
 C:Species: Mus musculus domesticus (western European house mouse)
 C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: A38211
 R:Morrison, M.E.; Racanietello, V.R.
 J. Virol. 66, 2807-2813, 1992
 A>Title: Molecular cloning and expression of a murine homolog of the human poliovirus re
 A:Reference number: A38211; MUID:92219365
 A:Accession: A38211
 A:Molecule type: DNA
 A:Residues: 1-467 <MOR>
 A:Cross-references: GB:M80206; NID:q199785; PIDN:AAA9734.1; PID:q199786
 C:Superfamily: poliovirus receptor; immunoglobulin homology
 C:Keywords: duplication; glycoprotein; transmembrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-67/Product: poliovirus receptor homolog #status predicted <MAT>
 F:26-354/Domain: extracellular #status predicted <EXT>
 F:47-133/Domain: immunoglobulin homology <IMM>
 F:167-231/Domain: immunoglobulin homology <IMM2>
 F:267-322/Domain: immunoglobulin homology <IMM3>
 F:355-374/Domain: transmembrane #status predicted <TMN>
 F:375-467/Domain: intracellular #status predicted <INT>
 F:54-131,174-229,274-320/Disulfide bonds: #status predicted

F:128,138,315/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 9; DB 1; Length 467;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PLPLPLLLL 14
 |||||
 Db 17 PLPLPLLLL 25

RESULT 11
 A53437
 poliovirus receptor mpvr - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
 C:Accession: A53437
 R:Aoki, J.; Koike, S.; Ise, I.; Sato-Yoshida, Y.; Nomoto, A.
 J. Biol. Chem. 269, 8431-8438, 1994
 A>Title: Amino acid residues on human poliovirus receptor involved in interaction wit
 A:Reference number: A53437; MUID:94179228
 A:Accession: A53437
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-530 <MOK>
 A:Cross-references: GB:D26107; NID:q475017; PIDN:BA05103.1; PID:9825507
 A:Experimental source: C57/BL6, Brain
 A:Note: Sequence extracted from NCBI backbone (NCBIN:146664, NCBI:146667)
 C:Superfamily: poliovirus receptor; immunoglobulin homology
 F:47-133/Domain: immunoglobulin homology <IMM>

Query Match 1.6%; Score 9; DB 2; Length 530;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PLPLPLLLL 14
 |||||
 Db 17 PLPLPLLLL 25

RESULT 12
 S44250
 integrin alpha-5 chain - mouse
 N:Alternate names: fibronectin receptor; vitronectin receptor alpha chain
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Sep-1999
 C:Accession: S44250; PL0103
 R:Morrisey, E.; Dutt, P.; Patel, V.
 submitted to the EMBL Data Library, April 1994
 A:Reference number: S44250
 A:Accession: S44250
 A:Molecule type: mRNA
 A:Residues: 1-1053 <MOR>
 A:Cross-references: EMBL:X79003; NID:q475910; PIDN:CAA5638.1; PID:q475911
 A:Experimental source: cell line murine erythroleukemia
 R:Holers, V.M.; Ruff, T.G.; Parks, D.L.; McDonald, J.A.; Ballard, L.L.; Brown, E.J.
 J. Exp. Med. 169, 1589-1605, 1989
 A>Title: Molecular cloning of a murine fibronectin receptor and its expression during
 for histocompatibility complex class II.
 A:Reference number: PL0103; MUID:89235580
 A:Accession: PL0103
 A:Molecule type: mRNA
 A:Residues: 645-1053 <HOL>
 A:Cross-references: GB:X15203; GB:Y00818; NID:q50984; PIDN:CAA33273.1; PID:q50985
 C:Comment: The receptor is a heterodimer consisting of alpha and beta chains.
 C:Keywords: glycoprotein; receptor; transmembrane protein
 F:997-1025/Domain: transmembrane #status predicted <TRM>
 F:1026-1053/Domain: intracellular #status predicted <INC>
 F:678,715,727,776,872,876/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 1.6%; Score 9; DB 2; Length 1053;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 LPLPLLL 14
 |||||
 DB 27 LPLPLLL 35

RESULT 13

S68700

HPRP beta-like tyrosine phosphatase precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: S68700

R:Kuramochi, S.; Matsuda, S.; Matsuda, Y.; Saitoh, T.; Ohnogi, M.; Yamamoto, T.

FEBS Lett. 378, 7-14, 1996

A:Title: Molecular cloning and characterization of Hprp, a murine receptor-type tyrosine

A:Reference number: S68700; MUID:96140699

A:Accession: S68700

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1238 <KUR>

A:Cross-references: GB:D45212; NID:q1208432; PIDN:BAA06146.1; PID:q1208433

C:Genetics:

A:Map position: 2E1-2

C:Superfamily: protein-tyrosine-phosphatase, receptor type J; fibronectin type III repe

C:Keywords: phosphoprotein

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1238/Product: HPRP beta-like tyrosine phosphatase #status predicted <MAT>

F:267-347/Domain: fibronectin type III repeat homology <3PR>

F:966-1188/Domain: protein-tyrosine-phosphatase homology <PTR>

F:1140/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1146/Binding site: substrate phosphate (Arg) #status predicted

Query Match

1.6%; Score 9; DB 2; Length 1238;

Best Local Similarity 100.0%; Pred. No. 4.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LPLPLLL 15
 |||||
 DB 21 LPLPLLL 29

RESULT 14

I48103

type VII collagen - Chinese hamster (fragment)

C:Species: Cricetus griseus (Chinese hamster)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Sep-1999

C:Accession: I48103

R:Greenspan, D.S.

Hum. Mol. Genet. 2, 273-278, 1993

A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous

A:Reference number: I48103; MUID:99271985

A:Accession: I48103

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1549 <RES>

A:Cross-references: GB:I06863; NID:q388624; PIDN:AAA36968.1; PID:q388625

C:Superfamily: unassigned collagens; animal kunitz-type proteinase inhibitor homology

F:1484-1536/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match

1.6%; Score 9; DB 2; Length 1549;

Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 384 CHPPVGGC 392
 |||||
 DB 1511 CHPPVGGC 1519

RESULT 15

T00093

hypothetical protein KIAA0467 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00093

R:Seki, N.; Ohira, M.; Nagase, T.; Ishikawa, K.; Miyajima, N.; Nakajima, D.; Nomura,

DNA Res. 4, 345-349, 1997

A:Title: Characterization of cDNA clones in size-fractionated cDNA libraries from hum

A:Reference number: Z14085; MUID:98116662

A:Accession: T00093

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2055 <SEK>

A:Cross-references: EMBL:AB007936; NID:q3413895; PIDN:BAA32312.1; PID:q3413896

A:Experimental source: brain

C:Genetics:

A:Map position: 1

A>Note: KIAA0467

Query Match

1.6%; Score 9; DB 2; Length 2055;

Best Local Similarity 100.0%; Pred. No. 6.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 WPPSSPGPP 171
 |||||
 DB 896 WPPSSPGPP 904

Search completed: February 26, 2002, 01:22:39
 Job time: 384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:21:26 ; Search time 162.6 Seconds
(without alignments)
19.589 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_75

Perfect score: 253

Sequence: 1 CPNGLSPNLMWDAOSTCERE.....DQCCAAAEKCCINVCGLHSC 43

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	253	100.0	216	22	Human membrane or
2	82.5	32.6	174	22	Mouse major epidid
3	82.5	32.6	174	22	Mouse epididymis-s
4	81	32.0	680	14	Sequence encoded b
5	81	33.0	680	19	Human KAL protein.
6	75	29.6	72	21	Polypeptide isolet
7	75	29.6	76	21	A mouse WDM-2 pro
8	74.5	29.4	125	12	HE4 epididymis-spe
9	74.5	29.4	125	12	Human HE4 protein.
10	74.5	29.4	125	22	Human epididymal p
11	72.5	28.7	124	21	Human secreted pro

12	72.5	28.7	124	22	AAV72647	Human epididymal p
13	72.5	28.7	137	21	AAV56783	Human prostate can
14	70	27.7	93	20	AAV36164	Human secreted pro
15	70	27.7	93	20	AAV36211	Human secreted pro
16	68	26.9	127	14	AAV30953	Rabbit whey acidic
17	67	26.5	49	22	AAV92379	Miscellaneous pept
18	67	26.5	61	11	AAV08146	Sodium potassium A
19	67	26.5	123	21	AAV70472	Human p53 target m
20	67	26.5	166	15	AAV50333	SPAI pre-sequence
21	67	26.5	189	15	AAV50334	SPAI peptide. Sus
22	66	26.1	45	17	AAV00661	Muteln #2 of secre
23	66	26.1	61	11	AAV08145	Generic sodium pot
24	66	26.1	98	17	AAV00666	Muteln #4 of secre
25	66	26.1	123	21	AAV90292	Human peptidase, H
26	65	25.7	45	17	AAV00660	Muteln #1 of secre
27	65	25.7	98	17	AAV00665	Muteln #3 of secre
28	64.5	25.5	37	21	AAV49445	Human Zdscl mature
29	64.5	25.5	37	22	AAE04335	Human mature disul
30	64.5	25.5	61	21	AAV68906	Amino acid sequenc
31	64.5	25.5	126	20	AAV98809	Mouse IMC carcino
32	64.5	25.5	131	20	AAV98908	Mouse IMC carcino
33	64.5	25.5	977	18	AAV36066	Soybean DNA-bindin
34	64	25.3	76	20	AAV28645	Human serine prote
35	64	25.3	111	21	AAV27652	Human protein PRO8
36	64	25.3	111	21	AAV66732	Membrane-bound pro
37	64	25.3	111	22	AAV65255	Human PRO844 (UNO5
38	64	25.3	2594	16	AAV14748	Human PRO844 (UNO5
39	64	25.3	5405	16	AAV14749	IgG-Fc binding pro
40	63	24.9	40	21	AAV49447	Porcine elafin FLE
41	63	24.9	40	22	AAE04337	Human FLE peptide.
42	63	24.9	57	11	AAV08217	Human leukocyte el
43	63	24.9	57	15	AAV48553	Elafin derivative
44	63	24.9	57	15	AAV48554	Elafin derivative
45	63	24.9	57	15	AAV48552	Elafin derivative

ALIGNMENTS

RESULT 1	
AAV88333	standard; Protein; 216 AA.
XX	
AC	AAV88333;
XX	
DT	23-MAY-2001 (first entry)
XX	
DE	Human membrane or secretory protein clone PSEC0040.
XX	
KW	Human; secretory protein; membrane protein; vaccine; gene therapy;
KW	Rheumatoid arthritis; diabetes.
XX	
OS	Homo sapiens.
XX	
PN	EP1067182-A2.
XX	
PD	10-JAN-2001.
XX	
PF	07-JUL-2000; 2000EP-0114090.
XX	
PR	08-JUL-1999; 99JP-0194179.
XX	
PR	11-JAN-2000; 2000JP-0118775.
XX	
PR	02-MAY-2000; 2000JP-0183766.
XX	
PA	(HELI-) HELIX RES INST.
PI	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX	
DR	WPI; 2001-093989/11.
XX	
DR	N-PSDB; AAF93760.
XX	
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -

XX Claim 1; SEQ ID 34; 609bp + CD ROM; English.

PS This invention relates to nucleic acid sequences AAF93744 - AAF93916
 XX which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419, included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.

XX Sequence 216 AA:

Query Match

Best Local Similarity 100.0%; Score 253; DB 22; Length 216;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CPNOLSPNLMVDAOSTCERCSDODCAAEKCCINVCGLHSC 43
 Db 33 cpnolspnmlwvdagstcercsdodcaaeckcincvclhsc 75

RESULT 2

AAE05360 standard; Protein; 174 AA.

AC AAE05360;

DT 12-SEP-2001 (first entry)

DE Mouse major epididymis-specific protein E4.

KW Mouse; cytosolic; antiinflammatory; immunoregulatory; tissue integrity;
 KW wound healing; immune response; vaccine; cancer; asthma; allergy;
 KW cell trafficking; therapy; major epididymis-specific protein E4;
 KW epididymal protein BE-20; secreted protein.

OS Mus sp.

PN WO200148192-A1.

PD 05-JUL-2001.

PF 21-DEC-2000; 2000MO-NZ00256.

PR 23-DEC-1999; 99US-0171678.

PR 28-NOV-2000; 2000US-0724864.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PI Watson JD, Morison JG;

DR WPI; 2001-425665/45.

DR N-PSDB; AADI0127.

PT Novel isolated polypeptide useful to isolate corresponding interacting
 PT proteins or other compounds, to quantitatively determine levels of
 PT interacting proteins or other compounds, and as therapeutic target
 PS Claim 6; Page 87-88; 101pp; English.

CC The patent discloses novel polynucleotides and their corresponding
 CC proteins which play a major role in induction of growth, cell migration-
 CC and proliferation, cell-cell interaction and the differentiation of
 CC tissue-specific cells. These proteins are important in the maintenance
 CC of tissue integrity and thus are important in wound healing. They are
 CC useful in various assays to determine the biological activity, to raise
 CC antibodies, to isolate corresponding interacting proteins or other
 CC compounds, to quantify corresponding levels of interacting proteins or
 CC other compounds, and as therapeutic target in a whole range of disease
 CC states. Compositions comprising the novel proteins of the invention are
 CC useful for treating mammalian disorders. Polynucleotides of the invention
 CC are useful in genome and physical mapping, in positional cloning of
 CC genes, to tag or identify an organism or its reproductive material (as
 CC non-disruptive tags for marking organisms), and for the diagnosis and
 CC treatment of mammalian diseases which is the consequence of inappropriate
 CC expression of kinase genes. They are useful for promoting immune response
 CC as part of a vaccine or anti-cancer treatment, as target for cancer
 CC treatment, as immunoregulatory and anti-inflammatory molecule, as
 CC anti-cancer treatment, and as a target for development of an
 CC or diseases such as asthma and allergy. They are also useful to inhibit
 CC or enhance the activity of the soluble molecule that binds proteins of
 CC the invention, for tissue and neural regeneration, to promote or block
 CC cell trafficking, and as anti-inflammatory and/or vaccine adjuvant.
 CC The present sequence is major epididymis-specific protein E4 (epididymal
 CC protein BE-20), a secreted protein from mouse.

XX Sequence 174 AA:

Query Match

Best Local Similarity 32.6%; Score 82.5; DB 22; Length 174;
 Matches 16; Conservative 4; Mismatches 18; Indels 5; Gaps 1;

OY 1 CPNOLSPNLMVDAOSTCERCSDODCAAEKCCINVCGLHSC 43
 Db 132 cpsvdlpk1-----glcedqcydsgscgsmkccrncgckmac 169

RESULT 3

AAB69093 standard; Protein; 174 AA.

AC AAB69093;

DT 23-APR-2001 (first entry)

DE Mouse epididymis-specific protein (HE4) protein SEQ ID NO:25.

KW Mouse; HPRT; hypoxanthine phosphoribosyl transferase; EPO;
 KW erythropoietin; recombination.

OS Mus musculus.

PN JP2000325091-A.

PD 28-NOV-2000.

PF 17-MAR-2000; 2000JP-0081795.

PR 17-MAR-1999; 99JP-0071390.

PA (NISB) JAPAN TOBACCO INC.

DR WPI; 2001-185036/19.

DR N-PSDB; AAF59010.


```
PT Preparation of a gene targeting DNA for homologous recombination and  
Pr screening of a cDNA library -  
  
PS Claim 82; Page 44-45; 59pp; Japanese.  
XX  
XX The present invention describes a method for the preparation of  
CC targeting DNA for modifying part of a DNA sequence in an endogenous  
CC genome of a living body by homologous recombination. The method  
CC comprises: (a) DNA Cl containing a genome DNA fragment containing the  
CC DNA sequence T is cleaved with restriction enzyme E1 at the restriction  
CC enzyme-recognition site X, which can be cleaved by the restriction  
CC enzyme E1 present respectively outside of said DNA sequence T to  
CC prepare a genome DNA fragment L1 having cleaved ends A and B sequences at  
CC both ends and containing the DNA sequence T; (b) DNA Fragment L1 is  
CC intermolecularly connected by the cleaved end of the DNA sequences to  
CC prepare a cyclic genome DNA C1 containing the restriction enzyme-  
CC recognition site X and sequence Y; (c) reverse polymerase chain  
CC reaction (PCR) is carried out using two primer DNAs P1 and P2 to  
CC prepare a recombinant genome DNA fragment L2; (d) the end of L2 is  
CC digested by restriction enzymes E2 and E3 to prepare a recombinant  
CC genome DNA fragment L2' and it is connected to a linearised expression  
CC vector to prepare a cyclic genome DNA C2; and (e) C2 is cleaved by  
CC restriction enzyme E1 to give a recombinant genome DNA fragment I3 and  
CC it is recovered as a targeting DNA. The method is useful for preparing  
CC targeting DNA. The present sequence represents a mouse epididymus-  
CC specific protein (Hb4), which is used in an example from the present  
CC invention.
```

```
SQ Sequence 174 AA;  
XX  
XX
```

```
OY Query Match 32.6%; Score 82.5; DB 22; Length 174;  
Best Local Similarity 37.2%; Pred. No. 0.057;  
Db Matches 16; Conservative 4; Mismatches 18; Indels 5; Gaps 1;
```

```
Dy      1 CPNQLSPNLWMDAOSTCEHCRRDDCAAEKCCINVGJLHSC 43  
||: || | :| | :| |: || | | | :|  
Dy    132 cpsvdipkrl-----glsedqcvdgscgsnmkcrrngcgknac 169
```

```
RESULT          4  
ID     AAR34445  
AC     AAR34445 standard; Protein; 680 AA.  
XX  
XX     AAR34445:  
DT     09-AUG-1993 (first entry)  
DE     Sequence encoded by a cDNA clone contg. the ADMLX gene  
DE associated with x-linked Kallmann Syndrome (KS).  
XX  
XX     x-linked Kallmann syndrome; ADMLX gene; diagnosis; PCR.  
KW  
XX Homo sapiens.  
OS  
XX  
XX W09307267-A.  
PN  
PD 15-APR-1993.  
PF 09-OCT-1992; 92MO-FR00956.  
PR 09-OCT-1991; 91FR-0012451.  
PS  
PA (INSP ) INST PASTEUR.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Claivette J, Hardelin J, Legouis R, Levisiers J,  
PI Lutfailla G, Petit C;  
DR WP1: 1993-134456/16.  
XR N-PSSB; AAA040341.  
XX
```

```
Nucleic acid sequence of gene with x-linked Kallmann syndrome -
```

Pt		useful for diagnosing Kallmann syndrome by amplification to detect genetic anomalies
Pt		
xx		
ps		Disclosure; Fig 6; 60bp; French.
xx		
cc		The nucleic acid sequence is derived from the ADMX gene associated with KS (or Hypogonadotropic hypogonadism and anosmia).
cc		Oligonucleotide pairs which act as primers for specific amplification of the gene are used in amplification methods to detect genetic anomalies which cause KS. The primer pairs corresp. to the coding CC and non-coding regions of exon 1 of the ADMX gene and one pair each for the other 13 exons.
cc		
xx	SQ	Sequence 680 AA;
Dq		Query Match 32.0%; Score 81; DB 14; Length 680; Best Local Similarity 34.9%; Pred. No. 0.29;
Dd		Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;
Oy		1 CPNQLSPNLWVDAOSTCEKCRDPCAAREKCCTINCGLHSC 43 : : : : : Db 134 cp---apekasgfaaacvsccevdnecsgykckcngcg-hlc 172
RESULT	5	
AAM49015		
AAW49015		AAW49015 standard; protein: 680 AA.
AC		AAW49015;
XX		
DT		29-SEP-1998 (first entry)
XX		
DE		Human KAL protein.
XX		
KM		Human KAL protein; retinal injury; renal injury; neuronal injury; neural injury; fibronectin type III repeat; 32R1 peptide.
OS		Homo sapiens.
XX		
FH		Key location/Qualifiers
FT		Domain 181..285
FT		/note= "Fibronectin type III repeat"
FT		Peptide 254..285
FT		/note= "32R1 peptide; this peptide is claimed by the inventors under claim 2 in the specification"
FT		Domain 286..402
FT		/note= "Fibronectin type III repeat"
FT		Domain 403..540
FT		/note= "Fibronectin type III repeat"
FT		Domain 541..661
FT		/note= "Fibronectin type III repeat"
XX		
PN		WO9824898-A2.
XX		
PD		11-JUN-1998.
PE		05-DEC-1997; 97WO-EPO6806.
PR		06-DEC-1996; 96US-O761136.
PA		(CNRS) CNRS CENT NAT RECH SCI. (INSP) INST PASTEUR.
PI		Ardouin O, Hardelin J, Legouis R, Mazie J, Petit C; Rougou G, Sarailh C, Soussi-Yanicostas N;
DR		WPI; 1998-333318/29.
PT		Use of KAL protein - for the treatment of traumatic, infectious, metabolic or inherited nerve, spinal, retinal or renal injury or disorder
PT		
PT		

The present sequence represents a polypeptide sequence which is isolated from lymph node stromal cells of fms^{-/-} mice. The polynucleotides and their polypeptides are useful for treating an inflammatory disorder, disorder of immune system and cancer selected from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a viral disorder, in particular HIV infection and for modulating the growth of blood vessels. The polypeptides are useful for treating a tumour necrosis factor (TNF) mediated disorder, such as those selected from arthritis, inflammatory bowel disease, such as those selected

The present sequence represents mouse MDN-2, which has homology to TANGO-175. TANGO-175 is related to several proteins in the four disulphide core family. Human TANGO-175 has activities similar to that of anti-leukoprotease and MDN-1 and may therefore have a role similar to these proteins by inhibiting proteinases associated with metastasis. The protein may play a role in regulating inflammation and also in the growth of hematopoietic stem cells by neutralizing proteinases produced by bone marrow accessory cells. TANGO-175 is therefore useful in treatment and diagnosis of cancer, inflammation and hematopoietic disorders. Primers and probes, which hybridize to human TANGO-175 nucleic acid molecules and antibodies against human-TANGO-175 protein, are useful for detecting the presence of the nucleic acid molecule or protein in a sample. The proteins and nucleic acids can be used to screen drugs or compounds, which modulate TANGO-175 activity or expression, to detect genetic lesions and to modulate TANGO-175 activity.

```
XX Sequence 76 AA;
SQ
Query Match 29.6%; Score 75; DB 21; Length 76;
Best Local Similarity 33.3%; Pred. No. 0.18;
Matches 16; Conservative 4; Mismatches 22; Indels 6; Gaps 2;
OY 1 CPNQL-----PNLWVDAOSTCERCSRDDCAAAEKCCINVGHLSC 43
DB 25 cpkefekpgecpkspesvgicvdqsgdgscpqnkccscscg-hvc 71
RESULT 8
AAR13329
ID AAR13329 standard; Protein; 125 AA.
XX
AC AAR13329;
XX
DT 10-OCT-1991 (first entry)
XX
DE HE4 epididymis-specific polypeptide.
XX
KM Diagnosis; probe; infertility; immuno sterilisation; therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
FT 31..125
FT Protein /label= mat_protein
XX
PN EP440321-A.
XX
PD 07-AUG-1991.
XX
PF 29-JAN-1991; 91EP-0250021.
XX
PR 30-NOV-1990; 90DE-4038189.
PR 01-FEB-1990; 90DE-4002981.
XX
PA (IHFH-) IHF INST HORMON FOR.
XX
PI Ivell R, Kirchhoff C;
XX
DR WPI; 1991-232030/32.
DR N-PSDB; AAO13127.
XX
PT Epididymis-specific DNA sequences - and corresp. polypeptides and
PT antibodies; useful for diagnosis or treatment of male infertility
PT or immuno-sterilisation
XX
PS Disclosure; Page 16-17; 28pp; English.
XX
CC DNA sequences have been identified by constructing a cDNA library
CC from human epididymis RNA, screening with epididymis and testis
CC probes, and further screening with brain and liver probes.
CC The DNA mol. is useful in the cloning and expression of human
CC epididymis-specific polypeptides in pro- or eukaryotic host cells.
CC The polypeptides and antibodies to the peptides are useful for
CC diagnosis or therapy of male infertility and for
CC immunosterilisation of mammals.
XX
SQ Sequence 125 AA;
Query Match 29.4%; Score 74.5; DB 12; Length 125;
Best Local Similarity 33.3%; Pred. No. 0.32;
Matches 16; Conservative 2; Mismatches 25; Indels 5; Gaps 1;
OY 1 CPNQL-----SPNLWVDAOSTCERCSRDDCAAAEKCCINVGHLSC 43
DB 31 cpnql-----SPNLWVDAOSTCERCSRDDCAAAEKCCINVGHLSC 43
```

```
DB 73 cpndkegscpqvnlfpqglcrdqgcqvtcpqgmkcrcnsgcgvsc 120
RESULT 9
AAW81779
ID AAW81779 standard; Protein; 125 AA.
XX
AC AAW81779;
XX
DT 23-FEB-1999 (first entry)
XX
DE Human HE4 protein.
XX
KM HE4; epididymis-specific; diagnosis; male infertility; treatment;
KM sterility; immunosterilisation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= signal
FT 31..125
FT Protein /label= HE4
XX
PN EP878544-A1.
XX
PD 18-NOV-1998.
XX
PF 29-JAN-1991; 91EP-0250021.
XX
PR 30-NOV-1990; 90DE-4038189.
PR 01-FEB-1990; 90DE-4002981.
XX
PA (IHFH-) IHF INST HORMON & FORTPELANZUNGS.
XX
PI Ivell R, Kirchhoff C;
XX
DR WPI; 1998-585748/50.
DR N-PSDB; AAV64622.
XX
PT DNA encoding human epididymis polypeptides - useful for, e.g.
PT diagnosis of male infertility
XX
PS Disclosure; Page 16-17; 29pp; German.
XX
CC This sequence represents a novel human epididymis-specific protein, HE4.
CC This sequence may be used for cloning and for expression of human
CC epididymis-specific polypeptides in prokaryotic or eukaryotic host
CC cells. Such proteins and antibodies generated from them may be used for
CC diagnosis of e.g. male infertility. The polypeptides and antibodies
CC may also be used for treatment of male infertility and for
CC immunosterilisation of mammals.
XX
SQ Sequence 125 AA;
Query Match 29.4%; Score 74.5; DB 19; Length 125;
Best Local Similarity 33.3%; Pred. No. 0.32;
Matches 16; Conservative 2; Mismatches 25; Indels 5; Gaps 1;
OY 1 CPNQL-----SPNLWVDAOSTCERCSRDDCAAAEKCCINVGHLSC 43
DB 73 cpndkegscpqvnlfpqglcrdqgcqvtcpqgmkcrcnsgcgvsc 120
RESULT 10
AAV72648
ID AAV72648 standard; Protein; 125 AA.
XX
AC AAV72648;
XX
DT 31-MAY-2001 (first entry)
XX
```

DE		Human epididymal protein HE4 #2.
KW	Human; epididymal protein; HE4; drug screening; therapy; inflammation;	
KW	renal disease; immunoglobulin A nephropathy; IgAN; anti-inflammatory;	
KW	necrotizing crescentic glomerulonephritis; NCGN; minimal change disease;	
KW	noninfectious systemic inflammatory response; SIRS; glomerulosclerosis;	
KW	mesangial cell proliferation; hypertension; proteinuria; nephrotropic;	
XX	haematuria.	
OS	Homo sapiens.	
PX		
PN	WO200116354-A1.	
XX		
PD	08-MAR-2001.	
XX		
PF	05-SEP-2000; 2000WO-US24282.	
XX		
PR	02-SEP-1999; 99US-0152123.	
XX		
PA	(GENE-) GENE LOGIC INC.	
PI	Munger WE, Sun H, Falk RJ.	
PT	WPI: 2001-226697/23.	
DR	N-PDB; AAD02714.	
XX		
PT	Novel nucleic acids encoding human epididymal proteins useful for	
PT	screening modulators of human epididymal proteins that are useful for	
XX	treating inflammatory diseases and renal diseases -	
PS	Claim 14; Page 48-49; 55pp; English.	
CC	The present sequence is human epididymal protein HE4 identified by	
CC	Kirchoff et al. The invention relates to changes in gene expression	
CC	in inflammatory diseases and renal diseases such as immunoglobulin A	
CC	nephropathy (IgAN) and necrotizing crescentic glomerulonephritis	
CC	(NCGN). HE4 sequence identified in IgAN kidney tissue is useful for the	
CC	diagnosis of inflammatory and renal diseases and screening modulators	
CC	of HE4. Modulator of HE4 is useful for treating inflammation or a renal	
CC	disease, such as IgAN, NCGN, minimal change disease and noninfectious	
CC	systemic inflammatory responses (SIRS). Modulator of HE4 is also useful	
CC	for modulating mesangial cell proliferation, hypertension,	
CC	glomerulosclerosis, haematuria and/or proteinuria.	
SQ		
Sequence	125 AA:	
Query Match	29.4%; Score 74.5; DB 22; Length 125;	
Best Local Similarity	33.3%; Pred. No. 0.32;	
Matches 16; Conservative	2; Mismatches 25; Indels 5; Gaps 1;	
OY	1 CPNQL-----SPNLVWDASTERCESRDQCAAEKCCINVCGLHSC 43	
Db	73 cpndkescgcpynlnfpqlglcrdqgydltqpgpqmccrncgcgvsc 120	
RESULT 11		
AAG03783		
ID	AAG03783 standard; Protein: 124 AA.	
AC	AAG03783;	
XX		
DT	06-OCT-2000 (first entry)	
XX		
DE	Human secreted protein, SF0 ID NO: 7864.	
XX		
KM	Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
KW	gene therapy; chromosome mapping.	
XX		
OS	Homo sapiens.	
PX		
PN	EPI033401-A2.	
XX		

```

PD 06-SEP-2000.
PE 21-FEB-2000; 2000EP-0200610.
PF XX
PR 26-FEB-1999; 99US-0122487.
XX
XX (GENSET ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX MPI; 2000-500381/45.
DR
XX N-PSDB; AAC03789.
XX
PS New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS
XX Claim 13; SEQ ID 7864; 71np + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 124 AA:
SQ
Query Match 28.7%; Score 72.5; DB 21; Length 124;
Best Local Similarity 34.9%; Pred No.0.53;
Matches 15; Conservative 2; Mismatches 23; Indels 3; Gaps 1;
QY 1 CENQLSPNLMVDAQSTCERECRSRDQCAAEKCCINVCGLHSC 43
Db 80 cp---gvninfpq1g1ordqcvdsqpgmkccrncgskvsc 119
RESULT 12
AAV72647
ID AAV72647 standard; Protein; 124 AA.
XX
XX AAV72647;
XX
XX 31-MAY-2001 (first entry)
XX
XX Human epididymal protein HE4 #1.
DE
XX Human; epididymal protein; HE4; drug screening; therapy; inflammation;
XX renal disease; immunoglobulin A nephropathy; IgAN; antiinflammatory;
XX necrotizing crescentic glomerulonephritis; NCN; minimal change disease;
XX noninfectious systemic inflammatory response; SIRS; glomerulosclerosis;
XX mesangial cell proliferation; hypertension; proteinuria; nephrotropic;
XX haematuria.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT 21..70 /label= WAP_I
FT /note= "Whey acidic protein (WAP)"
FT Domain 71..124 /label= WAP_II
FT /note= "Whey acidic protein (WAP)"
FT Misc-difference 71 /note= "this sequence contains Ser in place of Leu

```

```

FT      found in HE4 sequence identified by Kirchoff et al
FT      (See AAV72648)"
FT      Misc-difference 72..73
FT      /note= "This sequence contains Leu-Pro instead of
FT      Leu-Cys-Pro found in HE4 sequence identified by Kirchoff
FT      et al (See AAV72648)"
FT      Misc-difference 101
FT      /note= "This sequence contains Ser in place of Thr
FT      found in HE4 sequence identified by Kirchoff et al
FT      (See AAV72648)"
PN      WO200116354-A1.
PD      08-MAR-2001.
XX
XX      05-SEP-2000; 2000WO-US24282.
XX
XX      02-SEP-1999; 99US-0152123.
XX
XX      (GENE-) GENE LOGIC INC.
XX
XX      Munger WE, Sun H, Falk RJ;
XX      PI
XX      WPI: 2001-226697/23.
XX      N-PSDB; AAD02713.
XX
XX      Novel nucleic acids encoding human epididymal proteins, useful for
XX      screening modulators of human epididymal proteins that are useful for
XX      treating inflammatory diseases and renal diseases -
XX      PS
XX      Claim 14: Page 47; 55pp: English.
XX
XX      The present amino acid sequence is human epididymal protein HE4.
XX      CC      The HE4 sequence was identified in immunoglobulin A nephropathy (IgAN)
XX      kidney tissue. The invention relates to changes in gene expression in
XX      inflammatory diseases and renal diseases such as IgAN and necrotising
XX      crescentic glomerulonephritis (NCGN). HE4 sequence identified in
XX      immunoglobulin A nephropathy (IgAN) kidney tissue is useful for the
XX      diagnosis of inflammatory and renal diseases and screening modulators
XX      of HE4. Modulator of HE4 is useful for treating inflammation or a renal
XX      disease, such as IgAN, NCGN, minimal change disease and noninfectious
XX      systemic inflammatory responses (SIRS). Modulator of HE4 is also useful
XX      for modulating mesangial cell proliferation, hypertension,
XX      glomerulosclerosis, haematuria and/or proteinuria.
XX      CC
XX      Sequence 124 AA:
XX
XX      Query Match 28.7%; Score 72.5; DB 22; Length 124;
XX      Best Local Similarity 34.9%; Pred. No. 0.53;
XX      Matches 15; Conservative 2; Mismatches 23; Indels 3; Gaps 1.
OY      1 CPNQLSPILWVDAOSTCERCSRDDCAAEKCCINNGSLHSC 43
Db      80 cp---qvnifpqqlcrdqgvdgcpqymkccrncgqkvs 119
XX
XX      RESULT 13
XX      AAB56783
XX      AAB56783 standard; Protein: 137 AA.
XX
XX      AAB56783;
XX
XX      13-MAR-2001 (first entry)
XX
XX      Human prostate cancer antigen protein sequence SEQ ID NO:1361.
XX
XX      Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
XX      neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular;
XX      vulnervary; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
XX      antibacterial; gene therapy; neural; immune; reproductive; renal;
XX      gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX      wound; infectious disease.

```

XX	Homo sapiens.
OS	WO20005174-A1.
PN	21-SEP-2000.
XX	08-MAR-2000; 2000MO-US05988.
PD	12-MAR-1999; 99US-0124270.
XX	(HUMA-) HUMAN GENOME SCI INC.
PF	(ROSE/) ROSEN C A.
XX	Rosen CA, Ruben SM;
XX	WPI: 2000-587513/55.
DR	N-PSDB; AAF15986.
XX	prostate cancer associated gene sequences, referred to as prostate
PT	cancer antigens, useful for treatment, prevention, and diagnosis of
PT	disorders such as prostate cancer -
XX	Claim 11; Page 1788-1789; 233bp; English.
PS	AAF15566 to AAF16505 encode the human prostate cancer associated
CC	proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC	The prostate cancer antigens can have neuroprotective, cytosstatic,
CC	cardioactive, immunomodulatory, muscular, vlnerrary, gastrointestinal,
CC	neurotropic, antinfective, gynaecological and antibacterial activities,
CC	and can be used in gene therapy. The prostate cancer antigen
CC	polynucleotides may be used for detection of prostate cancer, chromosome
CC	identification, as chromosome markers, and for numerous other diagnostic
CC	or research purposes. The prostate cancer antigens may be used to treat
CC	disorders such as neural, immune, muscular, reproductive,
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC	disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC	AAB57303 represent sequences used in the exemplification of the present
CC	invention.
XX	Sequence 137 AA:
SQ	
Query Match	28.7%; Score 72.5; DB 21; Length 137;
Best Local Similarity	34.9%; Pred. No. 0.58;
Matches 15; Conservative	2; Mismatches 23; Indels 3; Gaps 1
QY	1 CGNQLSPNLMDVAQSTCERECRSRDDCAAAEKCCTIVCGLIHSC 43
DB	93 cp---gvnlnfpq1glcrtcdgcqvdsqcpgmkccrncgckvsc 132
RESULT 14	
AAAY36164	ID AAY36164 standard; protein; 93 AA.
AC	AAV36164;
DT	23-SEP-1999 (first entry)
DE	Human secreted protein #36.
KX	Secreted protein; human; cytosstatic; thrombotic; osteopathic; forensic;
KM	diagnostic; gene therapy; chromosome mapping; secretion vector.
OS	Homo sapiens.
PN	WO9925825-A2.
XX	27-MAY-1999.
PD	13-NOV-1998; 98WO-IB01862.
PF	
XX	

```

PR 04-SEP-1998; 98US-0099273.
PR 13-NOV-1997; 97US-0066677.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
XX
XX (GEST ) GENSET.
XX
XX Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
XX WPI: 1999-347472/29.
XX
XX N-PSDB; AAX97848.
XX
XX Extended cDNAs encoding secreted proteins
XX
XX Example 28; Page 246; 307pp; English.
XX
XX AAX36129-Y36222 represent novel human secreted proteins encoded by the
XX extended cDNA sequences represented in AAX97813-X97906. The proteins.
XX of the invention have cytosolic, thrombotic and osteopathic activity.
XX The extended cDNAs can be used to express secreted proteins or parts of
XX them or to obtain antibodies capable of binding to the secreted
XX proteins. They may also be used in diagnostic, forensic, gene therapy
XX and chromosome mapping procedures. Uses also include design of
XX expression vectors and secretion vectors.
XX
XX SQ Sequence 93 AA;

Query Match 27.7%; Score 70; DB 20; Length 93;
Best Local Similarity 30.6%; Pred. No. 0.77;
Matches 11; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

OY 5 LSPNLWVDAOSTCERECSDQDCAAEKCCINVCGL 40
   | | : | | : | | : | | : | | : | | :
Db 33 leppclsapencthlctmgedckgfcgssfcgi 68

RESULT 15
AAX36211
ID AAX36211 standard; protein; 93 AA.
XX
XX AAX36211;
XX
XX 23-SEP-1999 (first entry)
XX
XX Human secreted protein #83.
XX
XX Secreted protein; human; cytosolic; thrombotic; osteopathic; forensic;
XX diagnostic; gene therapy; chromosome mapping; secretion vector.
XX
XX Homo sapiens.
XX
XX WO925825-A2.
XX
XX 27-MAY-1999.
XX
XX 13-NOV-1998; 98WO-1B01862.
XX
XX 04-SEP-1998; 98US-0099273.
XX 13-NOV-1997; 97US-0066677.
XX 17-DEC-1997; 97US-0069957.
XX 09-FEB-1998; 98US-0074121.
XX 13-APR-1998; 98US-0081563.
XX 10-AUG-1998; 98US-0096116.
XX
XX (GEST ) GENSET.
XX
XX Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
XX WPI: 1999-347472/29.
XX
XX N-PSDB; AAX97895.
XX

```

```

XX
XX Extended cDNAs encoding secreted proteins
XX
XX Claim 7; Page 303; 307pp; English.
XX
XX AAX36129-Y36222 represent novel human secreted proteins encoded by the
XX extended cDNA sequences represented in AAX97813-X97906. The proteins
XX of the invention have cytosolic, thrombotic and osteopathic activity.
XX The extended cDNAs can be used to express secreted proteins or parts of
XX them or to obtain antibodies capable of binding to the secreted
XX proteins. They may also be used in diagnostic, forensic, gene therapy
XX and chromosome mapping procedures. Uses also include design of
XX expression vectors and secretion vectors.
XX
XX SQ Sequence 93 AA;

Query Match 27.7%; Score 70; DB 20; Length 93;
Best Local Similarity 30.6%; Pred. No. 0.77;
Matches 11; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

OY 5 LSPNLWVDAOSTCERECSDQDCAAEKCCINVCGL 40
   | | : | | : | | : | | : | | :
Db 33 leppclsapencthlctmgedckgfcgssfcgi 68

```

Search completed: February 26, 2002, 01:26:39
Job time: 313 sec

This page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:22:06 ; Search time 69.26 Seconds
(without alignments)
13,971 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_75

Perfect score: 253
Sequence: 1 CPNOLSPNLMWDASTCERE.....DQCCAAAKKCINVCGLHSC 43

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	32.4	42	2	US-08-761-248B-15
2	81	32.0	43	2	US-08-761-248B-13
3	81	32.0	680	1	US-08-211-430-2
4	81	32.0	680	3	US-08-761-136-1
5	77.5	30.6	44	2	US-08-761-248B-12
6	75	28.3	72	4	US-09-383-586-38
7	71.5	26.9	42	2	US-08-761-248B-9
8	68	26.9	127	4	US-08-162-146-3
9	68	26.9	127	4	US-09-314-127-3
10	67.5	26.7	51	2	US-08-761-248B-14
11	66.5	26.3	38	2	US-08-761-248B-11
12	64.5	25.5	37	4	US-09-326-039-5
13	64	25.3	2594	4	US-08-718-388-7
14	64	25.3	5405	4	US-08-718-388-9
15	63	24.9	40	4	US-09-326-039-7
16	63	24.9	42	2	US-08-761-248B-10
17	63	24.9	57	1	US-08-379-437-1
18	63	24.9	57	1	US-08-379-437-2
19	63	24.9	57	1	US-08-379-437-4
20	63	24.9	57	1	US-08-379-437-6
21	63	24.9	57	1	US-08-379-437-8
22	63	24.9	1104	2	US-08-327-832-5
23	63	24.9	1104	2	US-08-828-584-5
24	62	24.5	44	2	US-08-761-248B-8
25	62	24.5	107	2	US-07-963-538B-4
26	62	24.5	132	1	US-08-304-051-21
27	62	24.5	132	5	PCT-US95-11445-21

28	61	24.1	95	4	US-09-177-249-273	Sequence 273, App
29	61	24.1	689	4	US-09-177-249-2	Sequence 2, Appl1
30	61	24.1	689	4	US-09-061-769A-2	Sequence 2, Appl1
31	60	23.7	49	1	US-08-304-051-1	Sequence 1, Appl1
32	60	23.7	49	60	PCT-US95-11445-1	Sequence 1, Appl1
33	60	23.7	107	3	US-08-483-503A-4	Sequence 3, Appl1
34	59.5	23.5	49	2	US-07-963-538B-3	Sequence 3, Appl1
35	59.5	23.5	50	2	US-07-963-538B-2	Sequence 2, Appl1
36	59.5	23.5	53	2	US-07-963-538B-1	Sequence 1, Appl1
37	59.5	23.5	60	3	US-08-483-503A-1	Sequence 1, Appl1
38	58	22.9	1198	4	US-09-245-041-131	Sequence 131, App
39	58	22.9	1350	4	US-09-245-041-17	Sequence 17, App
40	58	22.9	1429	4	US-09-245-041-130	Sequence 130, App
41	58	22.9	2787	4	US-09-245-041-15	Sequence 15, Appl
42	58	22.9	4654	4	US-08-476-515A-84	Sequence 84, Appl
43	58	22.9	4655	4	US-08-652-877-84	Sequence 84, Appl
44	58	22.9	4655	4	US-08-652-877-86	Sequence 86, Appl
45	58	22.9	4655	4	US-08-652-877-90	Sequence 90, Appl

ALIGNMENTS

RESULT 1
US-08-761-248B-15
Sequence 15, Application US/08761248B
Patent No. 5958735
GENERAL INFORMATION:
APPLICANT: ROWLEY, DAVID R.
TITLE OF INVENTION: URGENTINAL SINUS DERIVED GROWTH
TITLE OF INVENTION: FACTOR NUCLEOTIDE AND AMINO ACID SEQUENCES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jenkins & Glitchist
STREET: 1100 Louisiana, Suite 1800
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77002
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/761,248B
FILING DATE: 06-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,348
FILING DATE: 07-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Turley, Charles P
REGISTRATION NUMBER: 35,723
REFERENCE/DOCKET NUMBER: 34012.6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713)9513310
TELEFAX: (713)9513314
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-761-248B-15

Query Match 32.4% ; Score 82; DB 2; Length 42;
Best Local Similarity 39.5% ; Pred. No. 0.0049;
Matches 17; Conservative 3; Mismatches 17; Indels 6; Gaps 2;

Wed Feb 27 08:23:06 200

OY 1 CPNOLSPNLTWYDAOSTCERCSRDODCAAEKCCINWGLHSC 43
DB 1 CPKMPRSI-----GTCVELCSGSDSCPNLQKCCSNGCG-HTC 37

RESULT 2
US-08-761-248B-13
Sequence 13
Patent No. 5948735
APPLICANT: Application US/08761248B
TITLE OF INVENTION: ROMELEY, DAVID R.
NUMBER OF INVENTION: PROCENTRAL SINUS DERIVED GROWTH
CORRESPONDENCE ADDRESS: 15
STREET: 1100 Genkens & Gilchrist
CITY: Houston
STATE: Louisiana, Suite 1800
COUNTRY: TX
ZIP: 77002
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: FASTSO DOS
CURRENT APPLICATION DATA:
FILING DATE: 06-DEC-1996
CLASSIFICATION: US/08/761,248B
PRIORITY APPLICATION: 435
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: TURTLES
REGISTRATION/DO NUMBER: 35,723
TELEPHONE: (713)9513310
TELEFAX: (713)9513314
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-761-248B-13

Query Match
Best Local Similarity 32.0%; Score 81; DB 2; Length 43;
Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;
OY 1 CPNOLSPNLTWYDAOSTCERCSRDODCAAEKCCINWGLHSC 43
DB 1 CP---APEKASGFMAACVCEVDNECSGVKCCSNGCG-HTC 39
RESULT 3
US-08-211-430-2
Sequence 2, Application US/08211430
Patent No. 5763166
APPLICANT:
TITLE OF INVENTION: NOCLETIC SEQUENCE OF THE GENE ASSOCIATED WITH
NUMBER OF INVENTION: X CHROMOSOME LINKED KALLMANN SYNDROME CORRESPONDING
COMPUTER READABLE FORM: 32
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

us-09-819-136-2_copy-33_75.tai

OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 1.0, Version #1.25 (PPO)
SEQUENCE CHARACTERISTICS:
LENGTH: 680 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL LINE: foetal brain cell
US-08-211-430-2

Query Match
Best Local Similarity 32.0%; Score 81; DB 1; Length 680;
Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;
OY 1 CPNOLSPNLTWYDAOSTCERCSRDODCAAEKCCINWGLHSC 43
DB 134 CP---APEKASGFMAACVCEVDNECSGVKCCSNGCG-HTC 172
RESULT 4
US-08-761-136-1
Sequence 1, Application US/08761136
Patent No. 6121231
APPLICANT:
APPLICANT: PETIT, CHRISTINE
APPLICANT: SOUSSY, JANTICOCRAS, NADIA
APPLICANT: SARAILIN, JEAN-PIERRE
APPLICANT: ROUGON, CATHERINE
APPLICANT: LEGOUTIS, RENAUD
APPLICANT: ARDOUIN, OLIVIER
TITLE OF INVENTION: USE OF KAL PROTEIN AND TREATMENT WITH
NUMBER OF INVENTION: THE KAL PROTEIN IN TREATMENT OF RETINAL, RENAL, NEU
CORRESPONDENCE ADDRESS:
ADDRESS: 1755 'C, SPIYAK, MCLELAND, MAIER & NEUSTADT,
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 1.0, Version #1.30
FILING DATE: 06-DEC-1996
CLASSIFICATION: US/08/761,136
ATTORNEY/AGENT INFORMATION:
NAME: ORION, NORMAN F.
REGISTRATION NUMBER: 24,618
TELEPHONE: 703-413-3000
TELEFAX: 703-413-3000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 680 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-761-136-1

Query Match 32.0%; Score 81; DB 3; Length 680;
Best Local Similarity 34.9%; Pred. No. 0.096;
Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;

OY 1 CPNQLSPNLWVDAQSTCERECSDQDCAAEKCCINVCGLHSC 43
DB 134 CP---APEKASGFPAACVESCSEVDNCGVKCKCCSNGCG-HFC 172

RESULT 5
US-08-761-248B-12
Sequence 12, Application US/08761248B
Patent No. 5958735

GENERAL INFORMATION:

APPLICANT: ROWLEY, DAVID R.
TITLE OF INVENTION: UROGENITAL SINUS DERIVED GROWTH
TITLE OF INVENTION: FACTOR NUCLEOTIDE AND AMINO ACID SEQUENCES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jenkins & Gilchrist
STREET: 1100 Louisiana, Suite 1800
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77002
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/761,248B
FILING DATE: 06-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,348
FILING DATE: 07-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Turley, Charles P
REGISTRATION NUMBER: 35,723
REFERENCE/DOCKET NUMBER: 34012.6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713)9513310
TELEFAX: (713)9513314
TELEX:

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-761-248B-12

Query Match 30.6%; Score 77.5; DB 2; Length 44;
Best Local Similarity 33.3%; Pred. No. 0.016;
Matches 13; Conservative 3; Mismatches 16; Indels 7; Gaps 1;

OY 1 CPNQLSPNLWVDAQSTCERECSDQDCAAEKCCINVCGLHSC 39
DB 1 CPEKTSFGFI-----CLHGCSDBSDCKEGQCKCFDGC 32

RESULT 6
US-09-383-586-38
Sequence 38, Application US/09383586

Patent No. 6242419
GENERAL INFORMATION:
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Abernethy, Nevin
APPLICANT: Onrust, Rene
APPLICANT: Kumble, Anand
APPLICANT: Murlison, Greg

Query Match 29.6%; Score 75; DB 4; Length 72;
Best Local Similarity 33.3%; Pred. No. 0.05;
Matches 16; Conservative 4; Mismatches 22; Indels 6; Gaps 2;

OY 1 CPNQLS-----PULWVDAQSTCERECSDQDCAAEKCCINVCGLHSC 43
DB 21 CPEKFGACPKPSPESVIGICVDQSGDSCPGNMKCCSNGC-HVC 67

RESULT 7
US-08-761-248B-9
Sequence 9, Application US/08761248B
Patent No. 5958735

GENERAL INFORMATION:

APPLICANT: ROWLEY, DAVID R.
TITLE OF INVENTION: UROGENITAL SINUS DERIVED GROWTH
TITLE OF INVENTION: FACTOR NUCLEOTIDE AND AMINO ACID SEQUENCES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jenkins & Gilchrist
STREET: 1100 Louisiana, Suite 1800
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77002
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/761,248B
FILING DATE: 06-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,348
FILING DATE: 07-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Turley, Charles P
REGISTRATION NUMBER: 35,723
REFERENCE/DOCKET NUMBER: 34012.6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713)9513310
TELEFAX: (713)9513314
TELEX:

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

APPLICATION NUMBER: US/08/718,388
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0230-111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2594 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-388-7

Query Match 25.3%; Score 64; DB 4; Length 2594;
Best Local Similarity 35.3%; Pred. No. 28;
Matches 18; Conservative 5; Mismatches 16; Indels 12; Gaps 3;

QY 4 QLSPLNWDASTCERCSC-----RD-ODCAAEKCCINWGLHSC 43
Db 807 QLAPEGVWADLCCORCTCNGATHOYTCRDKQSCPAGEKCSVQ-NGLLGC 856

RESULT 14
US-08-718-388-9
Sequence 9, Application US/08718388
Patent No. 6271362
GENERAL INFORMATION:
APPLICANT: MORIKAWA, MINORU
APPLICANT: HARADA, NAOKI
TITLE OF INVENTION: GENE ENCODING Igg FC REGION-BINDING
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,388
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0230-111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5405 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-388-9

Query Match 25.3%; Score 64; DB 4; Length 5405;
Best Local Similarity 35.3%; Pred. No. 58;
Matches 18; Conservative 5; Mismatches 16; Indels 12; Gaps 3;

QY 4 QLSPLNWDASTCERCSC-----RD-ODCAAEKCCINWGLHSC 43
Db 807 QLAPEGVWADLCCORCTCNGATHOYTCRDKQSCPAGEKCSVQ-NGLLGC 856

RESULT 15
US-09-326-039-7
Sequence 7, Application US/09326039
Patent No. 6239254
GENERAL INFORMATION:
APPLICANT: Conklin, Darrell
TITLE OF INVENTION: Disulfide Core Polypeptides
FILE REFERENCE: 98-13
CURRENT APPLICATION NUMBER: US/09/326,039
CURRENT FILING DATE: 1999-06-04
EARLIER APPLICATION NUMBER: 60/088,136
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 40
TYPE: PRT
ORGANISM: Homo sapiens
US-09-326-039-7

Query Match 24.9%; Score 63; DB 4; Length 40;
Best Local Similarity 45.0%; Pred. No. 0.62;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 21 CSRDDCAAEKCCINWGL 40
Db 15 CLKDTCPCGIRKCCGSCGM 34

Search completed: February 26, 2002, 01:28:02
Job time: 356 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:11 ; Search time 144.8 Seconds
(without alignments)
43.437 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_75

Perfect score: 253

Sequence: 1 CPNOLSPNLWVDAOSTCERE.....DDCAAAEKCCINVCGLHSC 43

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP_invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organella:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.5	35.4	262	13	Q98988
2	86	34.0	178	13	Q91AR5
3	85.5	33.8	633	13	Q91AR3
4	85	33.6	2150	5	O44131
5	83	32.8	675	13	Q9PSH7
6	82.5	32.6	174	11	Q9DAU7
7	82	32.4	674	13	Q90369
8	81	32.0	1511	5	Q9VB21
9	81	32.0	2174	5	Q9GQRO
10	81	32.0	3060	5	Q9VAV4
11	80.5	31.8	121	13	Q91450
12	80.5	31.8	225	11	Q99M36
13	79.5	31.4	131	5	Q23053
14	76.5	30.2	652	13	Q91AR4
15	73.5	29.1	1653	5	Q9VIT9
16	73.5	29.1	3680	5	Q9VR08
17	72.5	28.9	96	6	O62652
18	72.5	28.7	859	13	Q98UH9
19	72	28.5	3198	5	Q908G8

20	71	28.1	110	6	Q9XSV4	Q9XSV4	canis fam1
21	70.5	27.9	224	4	Q9BOP3	Q9BOP3	homo sapien
22	69	27.3	76	4	Q9H1F0	Q9H1F0	homo sapien
23	68.5	27.1	475	5	Q27087	Q27087	trichuris t
24	67.5	26.7	441	5	Q9W5X1	Q9W5X1	drosophila
25	67.5	26.7	903	5	O44397	O44397	trichuris t
26	66	26.1	207	6	Q9X543	Q9X543	phacocoeu
27	66	26.1	522	5	Q9VCC7	Q9VCC7	drosophila
28	65	25.7	114	6	Q9X546	Q9X546	sus scrofa
29	64.5	25.5	2475	4	Q9GEP7	Q9GEP7	caenorabd
30	64	25.3	104	4	Q9B831	Q9B831	homo sapien
31	64	25.3	654	5	Q17982	Q17982	caenorabd
32	64	25.3	1428	5	O44341	O44341	halictus ru
33	64	25.3	2843	4	Q9Y6R7	Q9Y6R7	homo sapien
34	63.5	25.1	181	6	P79389	P79389	sus scrofa
35	63	24.9	71	3	Q9P8B1	Q9P8B1	agarticus bl
36	63	24.9	114	4	O9Y5L8	O9Y5L8	homo sapien
37	63	24.9	818	11	Q9DBC8	Q9DBC8	mus musculi
38	63	24.9	949	5	P90956	P90956	caenorabd
39	63	24.9	1024	4	O9HX11	O9HX11	homo sapien
40	63	24.9	1106	5	O9Y025	O9Y025	drosophila
41	63	24.9	1114	11	Q90KW7	Q90KW7	mus musculi
42	62.5	24.7	180	6	O9X545	O9X545	hipopotamu
43	62.5	24.7	293	13	Q9DEX9	Q9DEX9	cyprinus ca
44	62.5	24.7	708	10	O9LGM8	O9LGM8	oryza sativ
45	62.5	24.7	806	5	O61677	O61677	lytechinus

ALIGNMENTS

RESULT 1

ID Q98988 PRELIMINARY: PRT: 262 AA.
AC Q98988;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OVULATORY PROTEIN-2 PRECURSOR.
OS Salvelinus fontinalis (Brook trout) (Brook char).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Procatantopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8038;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA Garczynski M.A., Goetz F.W.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U67854; AAB63598.1; .
DR HSSP: P1957; 1ELE.
DR InterPro: IPR002221; WAP.
DR Pfam: PF00095; WAP: 4.
DR PRINTS: PR00003; 4DISULPHCORE.
DR PRODOM: PD001224; WAP: 2.
DR SMART: SM00217; WAP: 3.
DR PROSITE: PS00317; 4_DISULFIDE_CORE: 5.
FT SIGNAL.
FT SIGNAL. 1 30 POTENTIAL.
FT CHAIN 31 262 OVULATORY PROTEIN-2.
SQ SEQUENCE 262 AA; 28200 MW; D2EC50FFAFECF807 CRC64;

Query Match

Best local Similarity 44.2%; Pred. No. 0.00014;
Matches 19; Conservative 1; Mismatches 16; Indels 7; Gaps 3;

QY 1 CPNOLSPNLWVDAOSTCERECSDDCAAAEKCCINVCGLHSC 43
DB 179 CRR-----RW--GVGTCAELGSDSDPCNDKCKHNGCG--HDC 214

RESULT 2

OGIAR5
ID OGIAR5 PRELIMINARY; PRT; 178 AA.
AC OGIAR5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE KAL1 (FRAGMENT).
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphae; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetracanthopterygii; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20054095; PubMed=10585565;
RA Ardouin O., Legouis R., Fasano L., David-Matine B., Korn H.,
RA Hardelin J.P., Petit C.;
RT "Characterization of the two zebrafish orthologues of the KAL-1 gene
underlying a chromosome-linked Kallmann syndrome.";
RL Mech. Dev. 90:89-94(2000).
DR EMBL; AF163309; AAF25778.1; -
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00095; wap; 1.
DR PRINTS; PR00003; 4DISULPHCORE.
DR ProDom; PD001224; WAP; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
FT NON_TER 1 178
FT SEQUENCE 178 AA; 1950 MW; BD8192BD7B4072AA CRC64;
SO

Query Match 34.0%; Score 86; DB 13; Length 178;
Best Local Similarity 39.5%; Pred. No. 0.00029;
Matches 17; Conservative 3; Mismatches 19; Indels 4; Gaps 2;
DB 1 CENQUSPINLVDAOSTCERCSRDQCAAAEKCCINVCGLHSC 43
28 CP--PPORATGFMAACVSCSSDCHCPSPRCSCNGCG-HTC 66

RESULT 3
OGIAR3 PRELIMINARY; PRT; 633 AA.
AC OGIAR3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE KAL1.2.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20054095; PubMed=10585565;
RA Ardouin O., Legouis R., Fasano L., David-Matine B., Korn H.,
RA Hardelin J.P., Petit C.;
RT "Characterization of the two zebrafish orthologues of the KAL-1 gene
underlying a chromosome-linked Kallmann syndrome.";
RL Mech. Dev. 90:89-94(2000).
DR EMBL; AF163311; AAF25780.1; -
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00095; wap; 1.
DR PRINTS; PR00003; 4DISULPHCORE.
DR ProDom; PD001224; WAP; 1.

DR SMART; SM00060; FN3; 4.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR SEQUENCE 633 AA; 70672 MW; DAC21803F8899E03 CRC64;
SO

Query Match 33.8%; Score 85.5; DB 13; Length 633;
Best Local Similarity 41.4%; Pred. No. 0.001;
Matches 12; Conservative 8; Mismatches 8; Indels 1; Gaps 1;
DB 15 STCERCSRDQCAAAEKCCINVCGLHSC 43
128 AACVSCAODRCSGVKCCSNGCG-HTC 155

RESULT 4
OGI431 PRELIMINARY; PRT; 2150 AA.
AC OGI431;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE C0869.2 PROTEIN.
GN C0869.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
OC Rhabdilitidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ahnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Fawcett A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Lattelle P.,
RA Lighthning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Snelson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thelley-Rieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Stellies L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA HSP; AF036687; AAB88311.1; -
DR EMBL; AF036687; AAB88311.1; -
DR HSP; PI2071; 2LFT.
DR InterPro; IPR002899; EB.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002223; Kunitz-BPTI.
DR InterPro; IPR000716; Thyroglobulin_1.
DR InterPro; IPR002221; WAP.
DR InterPro; IPR000822; ZnF-C2H2.
DR Pfam; PF00014; Kunitz-BPTI; 1.
DR Pfam; PF00086; thyroglobulin_1; 6.
DR Pfam; PF00095; wap; 9.
DR PRINTS; PR00003; 4DISULPHCORE.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD001224; WAP; 5.
DR SMART; SM00131; KU; 1.
DR SMART; SM00211; TY; 6.
DR SMART; SM00217; WAP; 9.

OS *Coturnix coturnix japonica* (Japanese quail).
OC *Eukaryota*; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC *Archosauvia*; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC *Coturnix*.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94010957; Pubmed=9406507;
RA Legouis R., Cohen-Salmon M., del Castillo I., Levlillers J., Capy L.,
RA Morinow J.P., Petit C.;
RT "Characterization of the chicken and quail homologues of the human
RT gene responsible for the x-linked Kallmann syndrome.";
RL Genomics 17:516-518(1993).
EMBL: L13976; AAA88500.1; -.
DR HSSP; P19957; 2REL.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002221; WAP.

DR SMART: SM00410; IG-like; 1.
 DR SMART: SM00131; KU; 3.
 DR SMART: SM00209; TSP1; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS50279; BPT1_KUNITZ_2; 3.
 DR PROSITE: PS50092; TSP1; 3.
 DR MATRIX protein: Serine protease inhibitor; signal.
 FT SIGNAL 1 26
 FT POTENTIAL.
 SO SEQUENCE 2174 AA: 231936 MW: 038F070952623120 CRC64:

Query Match
 Best Local Similarity 46.2%; Score 81; DB 5; Length 2174;
 Matches 18; Conservative 1; Mismatches 14; Indels 6; Gaps 2;

QY 1 CPNOLSPNLMVDAOSTCERCSDQCAAEKCCINVC 39
 DB 1874 CP-ALSN-----ASGCARCCYTDADCRGDNKCCSDCG 1906

RESULT 10
 Q9VAV4 PRELIMINARY: PRT; 3060 AA.
 AC Q9VAV4: Q9VAV3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG1540 PROTEIN.
 GN PPN OR CG1540 OR CG18436.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyroidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RX MEDLINE=20196006; PubMed=10731132;
 RC STRAIN-BERKELEY;
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratunga C., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Abdou A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
 RA Beeson K.Y., Bencs P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brodeur P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Jaitai M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RL "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AE003765; AAF56794.2; -.
 DR EMBL: AE003765; AAF56795.2; -.
 DR HSSP: P1211L; 2KNT.
 DR FLYBASE: FBgn0003137; Ppn.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003006; Ig_WHC.
 DR InterPro: IPR002223; Kunitz_BPT1.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00047; Ig; 3.
 DR Pfam: PF00014; Kunitz_BPT1; 12.
 DR Pfam: PF00090; TSP_1; 5.
 DR Pfam: PF00095; WAP; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00408; IGc2; 3.
 DR SMART: SM00131; KU; 12.
 DR SMART: SM00209; TSP1; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
 DR PROSITE: PS00280; BPT1_KUNITZ_1; 11.
 DR PROSITE: PS50279; BPT1_KUNITZ_2; 12.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS50092; TSP1; 3.
 KW Alternative splicing; Serine protease inhibitor.
 FT VARSPLIC 2803
 FT VARSPLIC 2844
 FT VARSPLIC 2854
 FT VARSPLIC 2855
 FT VARSPLIC 3060
 SO SEQUENCE 3060 AA: 331579 MW: ACA31D3E558C7C0 CRC64;

Query Match
 Best Local Similarity 46.2%; Score 81; DB 5; Length 3060;
 Matches 18; Conservative 1; Mismatches 14; Indels 6; Gaps 2;

QY 1 CPNOLSPNLMVDAOSTCERCSDQCAAEKCCINVC 39
 DB 2419 CP-ALSN-----ASGCARCCYTDADCRGDNKCCSDCG 2451

RESULT 11
 Q91450 PRELIMINARY: PRT; 121 AA.
 AC Q91450;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ANTI-EUKARYOTIC PRECURSOR.
 OS Salvelinus fontinalis (Brook trout) (Brook char).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Proactinopterygii; Salmoniformes; Salmonidae; Salvelinus.
 OX NCBI_TaxID=8038;
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RA Hsu S., Goetz F.W.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U03890; AAO3534.1; -.
 DR HSSP: P19957; 1FLC.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00095; WAP; 2.
 DR PRINTS: PR00003; 4DISULPHCORE.
 DR PRODOM: PD001224; WAP; 1.

DR SMART; SM00217; WAP; 2.
 DR PROSITE; PS00317; 4_DISULFIDE_CORE; 2.
 KW Signal.
 FT SIGNAL 1 29 POTENTIAL.
 CHAIN 30 121 ANTILEUKOPROTEINASE.
 SQ SEQUENCE 121 AA; 12734 MW; F407403777B4D3FE CRC64;

Query Match 31.8%; Score 80.5; DB 13; Length 121;
 Best Local Similarity 55.6%; Pred. No. 0.0017;
 Matches 15; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

OY 17 CERCSRDQCAAAEKCINVCGLHSC 43
 DB 48 CAELCSSDSDCPNDEKCHNCGC-HVC 73

RESULT 12
 099M36 PRELIMINARY; PRT; 225 AA.

AC 099M36; PRELIMINARY; PRT; 225 AA.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SIMILAR TO EXTRACELLULAR PROTEINASE INHIBITOR.

OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY TUMOR;

RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC002038; AA02038.1;
 DR EMBL; BC002038; AA02038.1;
 SQ SEQUENCE 225 AA; 24473 MW; BCC5254DE83B5919 CRC64;

Query Match

Best Local Similarity 31.8%; Score 80.5; DB 11; Length 225;
 Matches 13; Conservative 4; Mismatches 10; Indels 1; Gaps 1;

OY 16 TCERCSRDQCAAAEKCINVCGLHSC 43
 DB 194 TCDCRGTGSGSCSNMKCCSNCG-HAC 220

RESULT 13
 023053 PRELIMINARY; PRT; 131 AA.

AC 023053;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COSMID T27E4.

GN T27E4.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
 OC Rhabdilitidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Bonfield J., Burton J., Anderson K., Baynes C., Berks M.,
 RA Wilson R., Ainscough R., Connel M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Jones M., Kershaw J., Kirschen J., Laister N., Latteille P.,
 RA James D., Kershaw J., Kirschen J., Laister N., Latteille P.,
 RA Parsons J., Percy C., Riffen L., Roopra A., Mortimore B., O'Callaghan M.,
 RA Lightning J., Lloyd C., Momunray A., Saunders D., Shownkeen R.,
 RA Smalton N., Smith A., Sonnenhamer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Bradshaw H.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;

RA Waterston R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U64837; AAB04836.1;
 DR InterPro; IPR002221; WAP.
 DR SMART; SM00217; WAP; 1.
 SQ SEQUENCE 131 AA; 14725 MW; E3C5C81BD5EE72C CRC64;

Query Match 31.4%; Score 79.5; DB 5; Length 131;
 Best Local Similarity 35.4%; Pred. No. 0.0017;
 Matches 17; Conservative 6; Mismatches 20; Indels 5; Gaps 2;

OY 1 CPNOLSPNLMWDAOSTCE---REGSRDQC--AAEKCINVCGLHSC 43
 DB 57 CPSOLMAVFCOKTSTARKTVKPCXTDDCMFSNVQKCCDAGCGFNVG 104

RESULT 14
 091AR4 PRELIMINARY; PRT; 652 AA.

AC 091AR4;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE KAL1.1.

OS Brachydanio rerio (zebrafish) (zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.
 OX NCBI_Taxid=7955;

RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=20054095; PubMed=10585565;

RA Ardouin O., Legouis R., Fasano L., David-Watine B., Korn H.,

RA Hardelin J.P., Petit C.;
 RT "Characterization of the two zebrafish orthologues of the KAL-1 gene
 underlying x chromosome-linked Kallmann syndrome.";

RL Mech. Dev. 90:89-94(2000).
 DR EMBL; AF163310; AAF25779.1;
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002221; WAP.

DR Pfam; PF00041; fn3; 3.

DR Pfam; PF00095; WAP; 1.

DR PRINTS; PR00003; 4DISULPHORE.

DR PRODOM; PD001224; WAP; 1.

DR SMART; SM00060; FN3; 3.

DR SMART; SM00217; WAP; 1.

DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
 SQ SEQUENCE 652 AA; 72581 MW; 07D6D12D35FA1160 CRC64;

Query Match 30.2%; Score 76.5; DB 13; Length 652;
 Best Local Similarity 41.4%; Pred. No. 0.018;
 Matches 12; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

OY 15 STCERCSRDQCAAAEKCINVCGLHSC 43
 DB 136 AACVEGCEDEGCGSGKCCPGCG-HVC 163

RESULT 15

09YU9
ID 09YU9 PRELIMINARY; PRT: 1653 AA.
AC 09YU9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG10186 PROTEIN.
CN CG10186
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Splet E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003663; AAF53815.2; -
DR HSP: P10998; IYVD.
DR FLYbase: FBgn0032797; CG10186.
DR InterPro: IPR000923; Copper_blue1.
DR InterPro: IPR000566; Lipocln_cytfabp.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR InterPro: IPR002221; WAP.
DR Pfam: PF00084; sush1.14.
DR Pfam: PF00095; wap.1.
DR SMART: SM00032; CCP.17.
DR SMART: SM00217; WAP.1.
DR PROSITE: PS00196; COPPER_BLUE; UNKNOWN_1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1653 AA; 180864 MW; 174E7775C51FD60F CRC64;

Query Match 29.1%; Score 73.5; DB 5; Length 1653;
Best Local Similarity 36.0%; Pred. No. 0.1;

Matches 18; Conservative 8; Mismatches 9; Indels 15; Gaps 4;
QY 1 CPNQLSPNLMVDAOST-----CEREGSRDQCAAAEKCCI--NYCGIHSC 43
DB 48 CPRD-----EQQATILIGOKCLRKCKSSDDECKSKKKKCLDGVCG-NSC 89

Search completed: February 26, 2002, 01:49:14
Job time: 1263 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:56 ; Search time 45.78 Seconds
(without alignments)
34.438 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_75

Perfect score: 253
Sequence: 1 CPNQLSPMLNWDASTCERE.....DQDCAAEKCCINVCGLMSC 43

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83	32.8	676	1 KALM_CHICK	P33005 gallus gall
2	82	32.4	74	1 WDMN_RAT	P14730 rattus norv
3	81.5	32.2	123	1 EP4_RABIT	Q28631 oryctolagus
4	81	32.0	680	1 KALM_HUMAN	P23352 homo sapien
5	80.5	31.8	74	1 WDMN_MOUSE	Q62477 mus musculu
6	78.5	31.0	132	1 WAP_PIG	Q46655 sus scrofa
7	77.5	30.8	134	1 EPPI_MOUSE	Q9da01 mus musculu
8	77.5	30.6	110	1 IBP_CANCRA	P00933 caretta car
9	75.5	29.8	124	1 EP4_CANCRA	Q28894 canis fami
10	74.5	29.4	125	1 EP4_HUMAN	Q14508 homo sapien
11	74	29.2	117	1 WAP_CAMDR	P09837 camelus dro
12	71.5	28.3	55	1 CALD_CAVPO	P22075 cavia porce
13	68	26.9	127	1 WAP_RABIT	P09412 oryctolagus
14	67.5	26.7	137	1 WAP_RAT	P01174 rattus norv
15	67	26.5	187	1 SPAL_PIG	P16225 sus scrofa
16	66	26.1	191	1 WAP_MOUSE	Q9n018 macropus eu
17	65	25.7	1295	1 GLP1_CAEEL	P13508 caenorhabdi
18	64.5	25.5	131	1 ELAI_MOUSE	P97430 mus musculu
19	63	24.9	117	1 ELAI_HUMAN	P19937 homo sapien
20	63	24.9	144	1 WAP_PIG	Q29126 sus scrofa
21	63	24.9	1104	1 NEFL_HUMAN	Q12968 homo sapien
22	63	24.9	1106	1 STC_DROME	P40798 drosophila
23	62	24.5	28	1 ITRA_HUMAN	P30709 momordica c
24	62	24.5	132	1 ALKI_HUMAN	P03973 homo sapien
25	62	24.5	133	1 EPP1_MOUSE	Q95925 homo sapien
26	61	24.1	28	1 ITRA_MOUSE	P10295 momordica c
27	61	24.1	177	1 VP23_ASFB7	P23173 african swi
28	60.5	23.9	134	1 WAP_MOUSE	P01173 mus musculu
29	59.5	23.5	186	1 DHML_METFL	Q50425 methylodact
30	59	23.3	276	1 THIG_MOUSE	Q08710 mus musculu
31	58	22.9	28	1 IELI_MOUSE	P10296 momordica c
32	58	22.9	167	1 ELAI_PIG	Q29125 sus scrofa
33	58	22.9	237	1 IPDE_DICDI	P22549 dictyostell

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	KALM_CHICK	83	32.8	676	1 KALM_CHICK	P33005 gallus gall
2	KALM_CHICK	82	32.4	74	1 WDMN_RAT	P14730 rattus norv
3	KALM_CHICK	81.5	32.2	123	1 EP4_RABIT	Q28631 oryctolagus
4	KALM_CHICK	81	32.0	680	1 KALM_HUMAN	P23352 homo sapien
5	KALM_CHICK	80.5	31.8	74	1 WDMN_MOUSE	Q62477 mus musculu
6	KALM_CHICK	78.5	31.0	132	1 WAP_PIG	Q46655 sus scrofa
7	KALM_CHICK	77.5	30.8	134	1 EPPI_MOUSE	Q9da01 mus musculu
8	KALM_CHICK	77.5	30.6	110	1 IBP_CANCRA	P00933 caretta car
9	KALM_CHICK	75.5	29.8	124	1 EP4_CANCRA	Q28894 canis fami
10	KALM_CHICK	74.5	29.4	125	1 EP4_HUMAN	Q14508 homo sapien
11	KALM_CHICK	74	29.2	117	1 WAP_CAMDR	P09837 camelus dro
12	KALM_CHICK	71.5	28.3	55	1 CALD_CAVPO	P22075 cavia porce
13	KALM_CHICK	68	26.9	127	1 WAP_RABIT	P09412 oryctolagus
14	KALM_CHICK	67.5	26.7	137	1 WAP_RAT	P01174 rattus norv
15	KALM_CHICK	67	26.5	187	1 SPAL_PIG	P16225 sus scrofa
16	KALM_CHICK	66	26.1	191	1 WAP_MOUSE	Q9n018 macropus eu
17	KALM_CHICK	65	25.7	1295	1 GLP1_CAEEL	P13508 caenorhabdi
18	KALM_CHICK	64.5	25.5	131	1 ELAI_MOUSE	P97430 mus musculu
19	KALM_CHICK	63	24.9	117	1 ELAI_HUMAN	P19937 homo sapien
20	KALM_CHICK	63	24.9	144	1 WAP_PIG	Q29126 sus scrofa
21	KALM_CHICK	63	24.9	1104	1 NEFL_HUMAN	Q12968 homo sapien
22	KALM_CHICK	63	24.9	1106	1 STC_DROME	P40798 drosophila
23	KALM_CHICK	62	24.5	28	1 ITRA_HUMAN	P30709 momordica c
24	KALM_CHICK	62	24.5	132	1 ALKI_HUMAN	P03973 homo sapien
25	KALM_CHICK	62	24.5	133	1 EPP1_MOUSE	Q95925 homo sapien
26	KALM_CHICK	61	24.1	28	1 ITRA_MOUSE	P10295 momordica c
27	KALM_CHICK	61	24.1	177	1 VP23_ASFB7	P23173 african swi
28	KALM_CHICK	60.5	23.9	134	1 WAP_MOUSE	P01173 mus musculu
29	KALM_CHICK	59.5	23.5	186	1 DHML_METFL	Q50425 methylodact
30	KALM_CHICK	59	23.3	276	1 THIG_MOUSE	Q08710 mus musculu
31	KALM_CHICK	58	22.9	28	1 IELI_MOUSE	P10296 momordica c
32	KALM_CHICK	58	22.9	167	1 ELAI_PIG	Q29125 sus scrofa
33	KALM_CHICK	58	22.9	237	1 IPDE_DICDI	P22549 dictyostell

FT DOMAIN 537 657 FIBONECTIN TYPE-III 4.
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 549 549 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 676 AA; 76375 MW; 3FAC7ED82EA7E352 CRC64;

Query Match 32.8%; Score 83; DB 1; Length 676;
 Best Local Similarity 34.9%; Pred. No. 0.019;
 Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;

OY 1 CPNOLSPNLWVDAQSTCERCSDQCAAEKCCINCGHSC 43
 DB 129 CP--APEKASGFMAACVSEADSECSGVKCCSNGCG-HTC 167

RESULT 2

WDMN_RAT ID WDMN_RAT STANDARD; PRT; 74 AA.

AC P14730;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE WDMN1 PROTEIN PRECURSOR (FRAGMENT).

OS WDMN1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast carcinoma;

RC MEDLINE=8310901; PubMed=3136918;

RA Dear T.N., Ramsdew I.A., Kefford R.F.;

RT "Differential expression of a novel gene, WDMN1, in nonmetastatic rat mammary adenocarcinoma cells."

RT Cancer Res. 48:5203-5209(1988).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Thymus;

RC MEDLINE=91207400; PubMed=2018519;

RA Dear T.N., Kefford R.F.;

RT "The WDMN1 gene product is a novel member of the 'four-disulphide core' family of proteins."

RT Biochem. Biophys. Res. Commun. 176:247-254(1991).

CC -1- FUNCTION: INVOLVED IN THE METASTATIC POTENTIAL OF ADENOCARCINOMAS IN RAT. COULD HAVE PROTEINASE INHIBITING CAPACITY.

CC -1- TISSUE SPECIFICITY: SPLEEN, KIDNEY, BRAIN, LIVER, LUNG, AND CELL LINE RAT-2.

CC -1- SIMILARITY: CONTAINS 1 WAP-TYPE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X13309; CAA31688.1; ALT_INIT.

DR PIR: JH0390; JH0390.

DR HSSP: P19957; 2REL.

DR InterPro: IPR002221; WAP.

DR Pfam: PF00095; WAP; 1.

DR ProDom: PD001224; WAP; 1.

DR SMART: SM00217; WAP; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.

FT CHAIN 24 74 WDMN1 PROTEIN.
 FT DOMAIN 29 73 WAP.
 SQ SEQUENCE 74 AA; 7740 MW; 51802C70CD4F0521 CRC64;

Query Match 32.4%; Score 82; DB 1; Length 74;
 Best Local Similarity 39.5%; Pred. No. 0.0043;
 Matches 17; Conservative 3; Mismatches 17; Indels 6; Gaps 2;

OY 1 CPNOLSPNLWVDAQSTCERCSDQCAAEKCCINCGHSC 43
 DB 33 CPKNPPRSI-----GTCVELCSGDSQCPNIGKCCSNGCG-HVC 69

RESULT 3

EP4_RABIT ID EP4_RABIT STANDARD; PRT; 123 AA.

AC 028631;

DT 15-JUL-1998 (Rel. 36, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MAJOR EPIDIDYMS-SPECIFIC PROTEIN E4 PRECURSOR (EPIDIDYMAL PROTEIN BE-20).

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A. AND REVISION TO 29.

RC TISSUE=Epididymis;

RC MEDLINE=99201788; PubMed=10101572;

RA Fan H.-Y., Miao S.-Y., Wang L.-F., Koide S.S.;

RT "Expression and characterization of an epididymis-specific gene."

RT Arch. Androl. 42:63-69(1999).

RL [2]

RP SEQUENCE OF 29-123 FROM N.A. AND SEQUENCE OF 29-58.

RC TISSUE=Epididymis, and Epididymal fluid;

RC MEDLINE=97040973; PubMed=8886263;

RA Xu W.D., Wang L.-F., Miao S.-Y., Zhao M., Fan H.-Y., Zong S.D.,

RA Wu Y.W., Shi X.Q., Koide S.S.;

RT "Identification of a rabbit epididymal protein gene."

RT Arch. Androl. 37:135-141(1996).

CC -1- FUNCTION: POSSIBLE FUNCTION IN SPERM MATURATION.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- TISSUE SPECIFICITY: EPIDIDYMS.

CC -1- SIMILARITY: CONTAINS 2 WAP-TYPE DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U26725; AAA66525.2; --

DR HSSP: P19957; 1FLE.

DR InterPro: IPR002221; WAP.

DR Pfam: PF00095; WAP; 2.

DR ProDom: PD001224; WAP; 1.

DR SMART: SM00217; WAP; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.

DR PROSITE: PS00317; 4-DISULFIDE_CORE; 2.


```

RN      [2]
RP      SEQUENCE OF 17-132 FROM N.A.
RC      TISSUE=Mammary gland;
RA      Masei A.M., Hall A., Bell K.T.;
RT      "Cloning and characterisation of the porcine whey acidic protein.";
RL      Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RM      [3]
RX      3D-STRUCTURE MODELING OF 20-132
RD      MEDLINE=2014368; PubMed=1068016;
RA      Ranganathan S., Simpson K.J., Shaw D.C., Nicholas K.R.;
RT      "The whey acidic protein family: a new signature motif and three-
RL      dimensional structure by comparative modeling.";
RC      J. Mol. Graph. Model. 17:106-113(1999).
RL      - FUNCTION: COULD BE A PROTEASE INHIBITOR. MAY PLAY AN IMPORTANT
CC      ROLE IN MAMMARY GLAND DEVELOPMENT AND TISSUE REMODELLING.
CC      - SUBCELLULAR LOCATION: SECRETED.
CC      - TISSUE SPECIFICITY: MILK-SPECIFIC; MAJOR PROTEIN COMPONENT OF MILK
CC      WHEY.
CC      -----
CC      - SIMILARITY: CONTAINS 2 WAP-TYPE DOMAINS.
CC      -----
CC      This SWISS-PRO entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AI000221; CAA03950.1; -
DR      EMBL; AF034646; AAC72878.1; -.
DR      PDB; 1CTR; 15-APR-99.
DR      InterPro; IPR002221; WAP.
DR      Pfam; PF00095; wap; 2.
DR      PRINTS; PR00003; 4DISULPHCORE.
DR      ProDom; PD001224; WAP; 1.
DR      SMART; SM00217; WAP; 2.
DR      PROSITE; PS00317; 4_DISULFIDE_CORE; 2.
KM      MILK; Whey; Protease Inhibitor; Repeat; Signal; 3D-structure.
FT      SIGNAL 1 19
FT      CHAIN 20 132
FT      DOMAIN 28 73
FT      DOMAIN 79 126
FT      DOMAIN 36 39
FT      DISULFD 32 60
FT      DISULFD 43 65
FT      DISULFD 47 59
FT      DISULFD 53 69
FT      DISULFD 83 114
FT      DISULFD 95 118
FT      DISULFD 101 113
FT      DISULFD 107 122
FT      CONFLICT 39 40
FT      SEQUENCE 132 AA; 13956 MW; 7E7796493C0D98E0 CRC64;
QY      2 PNOISPNLWDAOSTGERECGRDQCAAEKCIYNVGL 40
Db      87 PAPIAPELCLE-----KNEGSRDQCGGNKKCCRSSCAM 120
Query Match 31.0%; Score 78.5; DB 1; Length 132;
Best Local Similarity 35.9%; Pred. NO. 0.017;
Matches 14; Conservative 7; Mismatches 13; Indels 5; Gaps 1;
RESULT 7
EPP1_MOUSE
ID EPP1_MOUSE STANDARD; PRT; 134 AA.
AC Q9DA01;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPPIN PRECURSOR.
Mus musculus (Mouse).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE-Testis;
 RX MEDLINE-21085660; Pubmed-11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Sessa T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 WAP-TYPE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AK006296; BAB24514.1; -;
 DR InterPro: IPR002221; WAP.
 DR ProSITE: PS002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR Pfam: PF00095; WAP; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 1.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
 KW Serine protease inhibitor; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 134 WAP.
 FT DOMAIN 29 73 BPTI/KUNITZ INHIBITOR.
 FT DISULFID 33 61 BY SIMILARITY.
 FT DISULFID 40 65 BY SIMILARITY.
 FT DISULFID 48 60 BY SIMILARITY.
 FT DISULFID 54 69 BY SIMILARITY.
 FT DISULFID 77 127 BY SIMILARITY.
 FT DISULFID 86 110 BY SIMILARITY.
 FT DISULFID 102 123 BY SIMILARITY.
 FT DISULFID 102 123 BY SIMILARITY.
 SQ SEQUENCE 134 AA; 15470 MW; DFE6B53D4D4C427F CRC64;

Query Match 30.8%; Score 78; DB 1; Length 134;
 Best Local Similarity 48.1%; Pred. No. 0.019;
 Matches 13; Conservative 4; Mismatches 6; Indels 4; Gaps 1;

OY 17 CERD-----CSRDDCAAAKCCINWCG 39
 11 : 1:1:1:1 1111: 11
 DB 40 CEHQERDLCTRRDRCPKKEKCCVFCNG 66

RESULT 8
 ID IBP_CARCR STANDARD; PRT; 110 AA.
 AC P00993;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHELONININ (BASIC PROTEASE INHIBITOR) (RPII).
 OS Caretta caretta (Loggerhead).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Testudines; Cryptodira; Chelonioidea; Chelonidae; Caretta.
 OC NCBI_TaxID-8467;
 RN [1]
 RP SEQUENCE.
 RC TISSUE-Egg white;
 RA Kato I., Tomimaga N.;
 RT "Trypsin-subtilisin inhibitor from red sea turtle eggwhite consists
 RT of two tandem domains -- one Kunitz -- one of a new family.";
 RL Fed. Proc. 38:832-832(1979).
 CC -1- FUNCTION: THE FIRST DOMAIN INHIBITS TRYPSIN; THE SECOND ONE
 CC INHIBITS SUBTILISIN.
 CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 WAP-TYPE DOMAIN.
 CC -1- CAUTION: AS THE PAPER ONLY INDICATES THE SPECIES AS "RED SEA
 CC TURTLE", THE SPECIES INDICATED HERE IS THEREFORE AN INFERENCE.
 DR PIR: A01224; TTYOR.
 DR HSSP: P00974; IBRB.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR ProSITE: PS002221; WAP.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR Pfam: PF00095; WAP; 1.
 DR PRINTS: PR00003; 4DISULPHORE.
 DR PRINTS: PR00759; BASICPTASE.
 DR PRODOM: PD001224; WAP; 1.
 DR SMART: SM00131; KU; 1.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
 KW Serine protease inhibitor.
 FT DOMAIN 8 58 BPTI/KUNITZ INHIBITOR.
 FT MOD_RES 63 105 WAP.
 FT DISULFID 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT DISULFID 8 58 BY SIMILARITY.
 FT DISULFID 17 41 BY SIMILARITY.
 FT DISULFID 33 54 BY SIMILARITY.
 FT DISULFID 67 92 BY SIMILARITY.
 FT DISULFID 76 97 BY SIMILARITY.
 FT DISULFID 80 93 BY SIMILARITY.
 FT DISULFID 86 101 BY SIMILARITY.
 FT ACT_SITE 18 19 REACTIVE BOND (TRYPSIN).
 SQ SEQUENCE 110 AA; 11916 MW; 269436243813418E CRC64;

Query Match 30.6%; Score 77.5; DB 1; Length 110;
 Best Local Similarity 33.3%; Pred. No. 0.019;
 Matches 13; Conservative 3; Mismatches 16; Indels 7; Gaps 1;

OY 1 CPNOLSPNLMVDAQSTCERCRCRSDQCAAAKCCINWCG 39
 11 : 1:1:1:1 1111: 11
 DB 67 CPRTSGPGT-----CLMGCDSDSDCKEGKCCFDCG 98

RESULT 9
 ID BP4_CANFA STANDARD; PRT; 124 AA.
 AC Q28894;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAJOR EPIDIDYMIS-SPECIFIC PROTEIN E4 PRECURSOR (CE4) (EPIDIDYMAL
 DE SECRETORY PROTEIN E4).

OS Canis familiaris (Dog).
 CC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedata; Canidae; Canis.
 ON NCBI_TaxId=9615;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Epididymis;
 RX MEDLINE=95263175; PubMed=7744511;
 RA Ellebroek R., Pera I., Hartung S., Iwell R.;
 RT "Gene expression in the dog epididymis: a model for human epididymal function."
 RL Int. J. Androl. 17:314-323(1994).
 RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=95263176; PubMed=7744512;
 RA Pera I., Iwell R., Kirchoff C.;
 RT "Regional variation of specific gene expression in the dog epididymis as revealed by in-situ transcript hybridization."
 RL Int. J. Androl. 17:324-330(1994).
 CC -1- FUNCTION: POSSIBLE FUNCTION IN SPERM MATURATION.
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: EPIDIDYMIS. HIGHEST LEVELS ARE FOUND IN THE CAUDAL AND PROXIMAL CAUDA REGIONS. LOWER LEVELS IN THE DISTAL CAUDA. NOT DETECTED IN THE EPIDIDYMAL DUCTS.
 CC -1- SIMILARITY: CONTAINS 2 MAP-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S77395; AAB34264.1; -
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF000095; WAP; 2.
 DR PRINTS: PR00003; 4DISULPHCORE.
 DR PRODOM: PD001224; WAP; 1.
 DR SMART: SM00217; WAP; 2.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE; 2.
 DR KW signal; Glycoprotein.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 124 MAJOR EPIDIDYMIS-SPECIFIC PROTEIN E4.
 FT DOMAIN 32 74 WAP 1.
 FT DOMAIN 76 124 WAP 2.
 FT DISULFID 36 62 BY SIMILARITY.
 FT DISULFID 45 66 BY SIMILARITY.
 FT DISULFID 49 61 BY SIMILARITY.
 FT DISULFID 55 70 BY SIMILARITY.
 FT DISULFID 80 110 BY SIMILARITY.
 FT DISULFID 93 114 BY SIMILARITY.
 FT DISULFID 97 109 BY SIMILARITY.
 FT DISULFID 103 119 BY SIMILARITY.
 FT CARBOHYD 44 44 N-LINKED (GLCNAG... (POTENTIAL).
 SQ SEQUENCE 124 AA; 12951 MW; 15A4F315BA13958C CRC64;.

Query Match 29.8%; Score 75.5; DB 1; Length 124;
 Best Local Similarity 37.8%; Pred. No. 0.035;
 Matches 17; Conservative 2; Mismatches 19; Indels 7; Gaps 2;
 OY 1 CP--NQLSPNIVDAQSTCERECGRDQDCAAEKCCINVCGLHSC 43
 DB 80 CPQVNTDFPOL-----GLCDQDCQVDSHCPLKCCYNGCGKVCSC 119

RESULT 10
 EP4_HUMAN
 ID EP4_HUMAN STANDARD; PRT; 125 AA.
 AC Q14508;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAJOR EPIDIDYMIS-SPECIFIC PROTEIN E4 PRECURSOR (HE4) (EPIDIDYMAL SECRETORY PROTEIN E4).
 GN HE4.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 ON NCBI_TaxId=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Epididymis;
 RX MEDLINE=92153963; PubMed=1686187;
 RA Kirchoff C., Habben L., Iwell R., Krull N.;
 RT "A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors."
 RL Biol. Reprod. 45:350-357(1991).
 CC -1- FUNCTION: POSSIBLE FUNCTION IN SPERM MATURATION.
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: EPITHELIAL CELLS OF THE EPIDIDYMAL DUCT, PREDOMINANTLY WITHIN THE DISTAL SECTIONS.
 CC -1- SIMILARITY: CONTAINS 2 MAP-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X63187; CAA44869.1; -
 DR EMBL: A18924; CAA01433.1; -
 DR HSSP: P19957; 1FDE.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF000095; WAP; 2.
 DR PRODOM: PD001224; WAP; 1.
 DR SMART: SM00217; WAP; 2.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE; 2.
 DR KW signal; Glycoprotein.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 125 MAJOR EPIDIDYMIS-SPECIFIC PROTEIN E4.
 FT DOMAIN 32 74 WAP 1.
 FT DOMAIN 77 125 WAP 2.
 FT DISULFID 36 62 BY SIMILARITY.
 FT DISULFID 45 66 BY SIMILARITY.
 FT DISULFID 49 61 BY SIMILARITY.
 FT DISULFID 55 70 BY SIMILARITY.
 FT DISULFID 81 111 BY SIMILARITY.
 FT DISULFID 94 115 BY SIMILARITY.
 FT DISULFID 98 110 BY SIMILARITY.
 FT DISULFID 104 120 BY SIMILARITY.
 FT CARBOHYD 44 44 N-LINKED (GLCNAG... (POTENTIAL).
 SQ SEQUENCE 125 AA; 13136 MW; B5BF5C08FF70E245 CRC64;.

Query Match 29.4%; Score 74.5; DB 1; Length 125;
 Best Local Similarity 33.3%; Pred. No. 0.046;
 Matches 16; Conservative 2; Mismatches 25; Indels 5; Gaps 1;
 OY 1 CPNQL-----SPNIVDAQSTCERECGRDQDCAAEKCCINVCGLHSC 43
 DB 73 CPNDKRGSCPVNINFPOLGLCDQDCQVDPQCGQKRCRNGCGKVCSC 120

RESULT 11
 WAP_CAMDR
 ID WAP_CAMDR STANDARD; PRT; 117 AA.
 AC P09837;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE WHEY ACIDIC PROTEIN (WAP).
 GN WAP.

FT CHAIN 20 127 WHEY ACIDIC PROTEIN.
 FT DOMAIN 28 68 WAP 1 (ATYPICAL).
 FT DOMAIN 74 121 WAP 2 (ATYPICAL).
 FT MOD_RES 237 237 PHOSPHORYLATION (PROBABLE).
 SQ SEQUENCE 127 AA; 13526 MW; 59DAD5F8FD2E2087 CRC64;

Query Match 26.98; Score 68; DB 1; Length 127;
 Best Local Similarity 31.1%; Pred. No. 0.25;
 Matches 14; Conservative 6; Mismatches 13; Indels 12; Gaps 3;

QY 1 CPNDLSPMLWYDA---OSTCER--ECGRDQDCAAEKCCINVCGL 40
 Db 78 CP-----WYQAPMLSQLCELSDCANDIECRGDKRCESRCAM 115

RESULT 14
 WAP_RAT
 ID WAP_RAT STANDARD; PRT; 137 AA.
 AC P01174;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE WHEY ACIDIC PROTEIN PRECURSOR (WHEY PHOSPHOPROTEIN) (WAP).
 GN WAP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85062841; PubMed=6095207;
 RA Campbell S.M., Rosen J.M., Hennighausen L.G., Strech-Yurk U.,
 RT Sippel A.E.;
 RT "Comparison of the whey acidic protein genes of the rat and mouse.";
 RT Nucleic Acids Res. 12:8685-8697(1984).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82274212; PubMed=6896749;
 RA Hennighausen L.G., Sippel A.E.;
 RT "Comparative sequence analysis of the mRNAs coding for mouse and rat
 RT whey protein.";
 RT Nucleic Acids Res. 10:3733-3744(1982).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82275050; PubMed=6955785;
 RA Dandekar A.M., Robinson E.A., Appella E., Qasba P.K.;
 RT "Complete sequence analysis of cDNA clones encoding rat whey
 RT phosphoprotein: homology to a protease inhibitor.";
 RT Proc. Natl. Acad. Sci. U.S.A. 79:3987-3991(1982).
 CC -1- FUNCTION: COULD BE A PROTEASE INHIBITOR.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: MILK-SPECIFIC; MAJOR PROTEIN COMPONENT OF MILK
 CC WHEY.
 CC -1- SIMILARITY: CONTAINS 2 WAP-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X01153; CAA25600.2; -;
 DR EMBL; X01154; CAA25600.2; JOINED.
 DR EMBL; X01155; CAA25600.2; JOINED.
 DR EMBL; X01156; CAA25600.2; JOINED.
 DR EMBL; J00802; AAA42347.1; -;
 DR EMBL; J00801; AAA42346.1; -;
 DR PIR; A01228; WYRT.
 DR PIR; A23879; A23879.
 DR HSSP; P19957; IFLE.

DR InterPro: IPR002221; WAP.
 DR Pfam: PF00095; Wap; 2.
 DR PRINTS: PR00003; 4DISULPHORE.
 DR ProDom: PD001224; WAP; 1.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4.DISULFIDE CORE; 2.
 KM Milk; Whey; Protease inhibitor; Repeat; Phosphorylation; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 137
 FT DOMAIN 30 73
 FT WAP 1 (ATYPICAL).
 FT DISULFID 34 61
 FT BY SIMILARITY.
 FT DISULFID 45 65
 FT BY SIMILARITY.
 FT DISULFID 48 60
 FT BY SIMILARITY.
 FT DISULFID 54 69
 FT BY SIMILARITY.
 FT DISULFID 83 115
 FT BY SIMILARITY.
 FT DISULFID 96 119
 FT BY SIMILARITY.
 FT DISULFID 102 114
 FT BY SIMILARITY.
 FT DISULFID 108 123
 FT BY SIMILARITY.
 FT MOD_RES 38 38
 FT BY SIMILARITY.
 FT MOD_RES 40 40
 FT BY SIMILARITY.
 FT MOD_RES 50 50
 FT BY SIMILARITY.
 FT MOD_RES 103 103
 FT BY SIMILARITY.
 FT CONFLICT 4 4
 FT S -> F (IN REF. 3).
 FT CONFLICT 35 35
 FT S -> P (IN REF. 2 AND 3).
 FT CONFLICT 39 39
 FT F -> S (IN REF. 2 AND 3).
 FT CONFLICT 47 47
 FT N -> K (IN REF. 2).
 FT CONFLICT 68 68
 FT S -> P (IN REF. 2).
 FT CONFLICT 99 99
 FT D -> G (IN REF. 2).
 FT CONFLICT 116 116
 FT K -> N (IN REF. 2 AND 3).
 FT CONFLICT 127 127
 FT E -> K (IN REF. 2).
 FT CONFLICT 129 129
 FT K -> E (IN REF. 3).
 FT CONFLICT 129 129
 FT K -> D (IN REF. 2).
 FT CONFLICT 134 134
 FT I -> V (IN REF. 3).
 SQ SEQUENCE 137 AA; 14827 MW; 1C2E8ADA9FD97949 CRC64;

Query Match 26.7%; Score 67.5; DB 1; Length 137;
 Best Local Similarity 42.4%; Pred. No. 0.3;
 Matches 14; Conservative 4; Mismatches 12; Indels 3; Gaps 2;

QY 13 AOSTCERE--CSRQDCAAEKCCINVCGLHSC 43
 Db 92 AAGCPKDPNCSDSDSGTMRCKCKNGC-IMSC 123

RESULT 15
 SPAL_PIG
 ID SPAL_PIG STANDARD; PRT; 187 AA.
 AC P16225;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SODIUM/POTASSIUM ATPASE INHIBITOR SPAL-2 PRECURSOR (WAP-2 PROTEIN).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-58.
 RC TISSUE=Duodenum;
 RX MEDLINE=95403443; PubMed=7673229;
 RA Kuroki J., Hosoya T., Itakura M., Hirose S., Tamechika I.,
 RA Yoshimoto T., Ghoneim M.A., Nara K., Kato A., Suzuki Y., Furukawa M.,
 RA Tachibana S.;
 RT "Cloning, characterization, and tissue distribution of porcine SPAL,
 RT a protein with a transglutaminase substrate domain and the WAP
 RT motif.";
 RT J. Biol. Chem. 270:22428-22433(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96215132; PubMed=8636131;
 RA Tamechika I., Itakura M., Saruta Y., Furukawa M., Kato A.,

RA Tachibana S., Hirose S.;
 RT "Accelerated evolution in inhibitor domains of porcine elafin family
 RL J. Biol. Chem. 271:7012-7018(1996).
 RN [3]
 RP SEQUENCE OF 127-187.
 RC TISSUE-DUODENUM;
 RX MEDLINE-90026425; PubMed-2553020;
 RA Araki K., Kuroki J., Ito O., Kuwada M., Tachibana S.;
 RT "Novel peptide inhibitor (SPAI) of Na⁺, K⁺-ATPase from porcine
 RL intestine."
 RN Biochem. Biophys. Res. Commun. 164:496-502(1989).
 RL [4]
 RP DISULFIDE BONDS;
 RX MEDLINE-91025070; PubMed-2171523;
 RA Araki K., Kuwada M., Ito O., Kuroki J., Tachibana S.;
 RT "Four disulfide bonds' allocation of Na⁺, K⁺-ATPase inhibitor
 RL (SPAI)."
 CC -1- FUNCTION: INHIBITS NA⁺, K⁺ ATPASE BY THE COMPETITIVE MODE
 AGAINST NA⁺.
 CC -1- TISSUE SPECIFICITY: SMALL INTESTINE > LARGE INTESTINE. THE PLASMA
 CONTAINS THE PRO-SPAI FORM CIRCULATING.
 CC -1- DOMAIN: THE REPETITIVE DOMAIN OF PRO-SPAI SERVES AS A SUBSTRATE
 FOR TRANSGLUTAMINASE.
 CC -1- PTM: THE SHORT FORM (AA 127-187) MAY BE AN ARTIFACT DUE TO THE
 STRONGLY ACIDIC CONDITIONS OF THE DUODENUM. THE PRO-SPAI FORM MAY
 BE THE NATIVE FORM.
 CC -1- SIMILARITY: CONTAINS 1 WAP-TYPE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D17753; BAA04602.1; -;
 DR EMBL; D17756; BAA04603.1; -;
 DR EMBL; D17754; BAA04603.1; JOINED.
 DR EMBL; D17755; BAA04603.1; JOINED.
 DR EMBL; D83667; BAA12037.1; -;
 DR EMBL; D50320; BAA08855.1; -;
 DR PIR; A33429; A33429.
 DR PIR; B33429; B33429.
 DR PIR; C33429; C33429.
 DR HSSP; P12071; 2ETI.
 DR InterPro: IPR002098; SVP_1.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00095; WAP_1.
 DR PRINTS; PRO0003; 4DISULPHCORE.
 DR ProDom; PD001224; WAP; 1.
 DR SMART; SM00217; WAP; 1.
 DR PROSITE; PS00313; SVP_1; 2.
 DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
 KM Signal; Repeat.
 FT SIGNAL 1 21
 FT PROPEP 22 126
 FT CHAIN 127 187
 FT REPEAT 62 85
 FT DOMAIN 34 129
 FT REPEAT 34 39
 FT REPEAT 34 39
 FT REPEAT 40 45
 FT REPEAT 46 51
 FT REPEAT 58 63
 FT REPEAT 64 69
 FT REPEAT 70 75
 FT REPEAT 76 81
 FT REPEAT 82 87
 FT REPEAT 88 93
 FT REPEAT 100 105
 FT REPEAT 106 111
 SODIUM/POTASSIUM ATPASE INHIBITOR SPAI-2.
 SVP-1 CLOTTING 1.
 14 X 6 AA APPROXIMATE TANDEM REPEATS.

FT REPEAT 112 117 12.
 FT REPEAT 118 123 13.
 FT REPEAT 124 129 14.
 FT DOMAIN 142 187 WAP.
 FT MOD_RES 22 22 PYRROLIDONE CARBOXYLIC ACID.
 FT DISULFID 146 175
 FT DISULFID 153 179
 FT DISULFID 162 174
 FT DISULFID 168 183
 FT VARIANT 127 138
 FT VARIANT 148 148
 FT VARIANT 156 156
 SQ SEQUENCE 187 AA; 20471 MW; 12FSBD0813AF5E27 CRC64;
 MISSING (IN SPAI-1).
 R -> G (IN SPAI-3).
 S -> G (IN SPAI-3).

Query Match 26.5%; Score 67; DB 1; Length 187;
 Best Local Similarity 44.0%; Pred. No. 0.44;
 Matches 11; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 15 STCERECRQDDCAAEKCCINVG 39
 DB 156 SNPSNKCWRDYPDCGVKKCCGFCG 180

Search completed: February 26, 2002, 01:46:36
 Job time: 1180 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:22:41 ; Search time 78.07 Seconds
(without alignments)
41.956 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_75

Perfect score: 253
Sequence: 1 CPNOLSPNLMVDAOSTCERE.....DQDCAAEKCCINVCGLHSC 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_68:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	33.6	2150	2 T32497	hypothetical prote
2	83	32.8	676	2 B47222	Kallmann syndrome
3	82	32.4	68	2 S07807	WDNMI protein - ra
4	82	32.4	674	2 A47222	Kallmann syndrome
5	81	32.0	679	2 A40351	adhesion-type prot
6	81	32.0	680	2 S17982	Kallmann syndrome
7	80.5	31.8	74	2 S61553	probable proteinase
8	79.5	31.4	131	2 T25924	hypothetical prote
9	77.5	30.6	110	1 T1770R	basic proteinase 1
10	75.5	29.8	124	2 I54768	epididymis-specific
11	74.5	29.4	125	2 S25454	epididymis-specific
12	74	29.2	117	2 A24178	Whey acidic protei
13	73	28.9	96	2 JE0252	trypsin-6 bovine
14	71.5	28.3	55	2 B35752	caltrin-like prote
15	68.5	27.1	478	2 S47040	gene T52 protein
16	68	26.9	127	2 S01286	whey acidic protei
17	67.5	26.7	137	1 WYR	Whey acidic protei
18	67	26.5	187	2 I46650	ATPase inhibitor S
19	66	26.1	61	2 C33420	ATPase inhibitor 3
20	66	26.1	207	2 JE0250	trypsin-1 - wart h
21	65	25.7	114	2 JE0255	trypsin-9 - pig
22	65	25.7	1295	2 A32901	glp1 protein precu
23	64.5	25.5	2195	2 T34264	hypothetical prote
24	64	25.3	654	2 T30136	hypothetical prote
25	64	25.3	1428	2 T08852	lustrin A - Califo
26	63.5	25.1	181	2 JE0253	trypsin-7 - pig
27	63	24.9	117	2 JH0614	elafin precursor -
28	63	24.9	949	2 T24294	hypothetical prote
29	63	24.9	1104	2 I38869	transcription fact

30	63	24.9	1106	2 T13938	gene shuttle craf
31	63	24.9	1106	2 T44598	hypothetical prote
32	62.5	24.7	180	2 JE0257	trypsin-11 - hippo
33	62	24.5	28	2 S20393	trypsin inhibitor
34	62	24.5	132	1 T1H0SP	antileukoproteinas
35	62	24.5	153	2 JE0256	trypsin-10 - colla
36	61.5	24.3	134	1 WYMS	whey acidic protei
37	61	24.1	28	2 JX0058	trypsin inhibitor
38	61	24.1	689	2 T52060	protein MEDPA [imp
39	60	23.7	27	2 JC2507	trypsin inhibitor
40	59.5	23.5	496	2 T33496	hypothetical prote
41	59.5	23.3	662	2 T23271	hypothetical prote
42	59	23.3	115	2 JE0254	trypsin-8 - pig
43	59	23.3	201	2 D71190	hypothetical prote
44	58.5	23.1	269	2 A85049	probable transpos
45	58.5	23.1	1142	2 T30272	hypothetical prote

ALIGNMENTS

RESULT 1
T32497
hypothetical protein C08G9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32497
R:Geisel, C.; Stellyes, L.
Submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid C08G9.
A:Reference number: 221179
A:Accession: T32497
A:Status: preliminary; translated from GB/EMBL/DDAT
A:Molecule type: DNA
A:Residues: 1-2150 <GET>
A:Cross-references: EMBL:AF036687; PIDN:AA88311.1; GSPDB:GN00022; CESP:C08G9.2
A:Experimental source: strain Bristol N2; clone C08G9
C:Genetics:
A:Gene: CESP:C08G9.2
A:Map position: 4
A:Map position: 4
A:Intons: 242/1; 306/1; 340/3; 485/1; 545/1; 736/2; 791/1; 829/1; 886/3; 942/1; 1079

Query Match 33.6%; Score 85; DB 2; Length 2150;
Best Local Similarity 36.6%; Pred. No. 0.13;
Matches 15; Conservative 8; Mismatches 10; Indels 8; Gaps 2;

OY 1 CPNOLSPNLMVDAOSTC--ERECSRDQDCAAEKCCINVC 39
DB 129 CPSPPTS-----NKSTCIDQKCCANDLDGCVKCCGCMAG 163

RESULT 2
B47222
Kallmann syndrome protein homolog KAL - chicken
C:Species: Gallus gallus (chicken)
C>Date: 24-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000
C:Accession: B47222; A47394; S36170
R:Legouis, R.; Cohen-Salmon, M.; del Castillo, I.; Levilliers, J.; Caby, L.; Morrow, Genomics 17, 516-518, 1993
A:Title: Characterization of the chicken and quail homologues of the human gene respo
A:Reference number: A47222; MIMD:94010957
A:Accession: B47222
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-676 <LEG>
A:Cross-references: GB:U12144; NID:q406510; PIDN:AA51435.1; PID:q406511
A>Note: sequence extracted from NCBI backbone (NCBIP:137996)
R:Legouis, R.; Llevre, C.A.; Lebouvier, M.; Lapointe, F.; Petit, C.
Proc. Natl. Acad. Sci. U.S.A. 90, 2461-2465, 1993
A:Title: Expression of the KAL gene in multiple neuronal sites during chicken develop
A:Reference number: A47394; MIMD:93211981
A:Accession: A47394

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 81-152, 'P', 154-237 <LE2>
 A:Experimental source: embryo, olfactory bulb
 R:Rugari, E.I.; Lutz, B.; Kuratani, S.C.; Wawersik, S.; Borsani, G.; Ballabio, A.; Eich
 A:Title: Sequence extracted from NCBI backbone (NCBI:128286, NCBI:128287)
 A:Reference number: 536170; MUID:93291868
 A:Accession: S36170; MUID:93291868
 A:Molecule type: mRNA
 A:Superfamily: MVR/5-528, 'H', 530-676 <MG>
 F:125-171/Domain: antileukoproteinase repeat homology

Query Match
 Best Local Similarity 32.8%; Score 83; DB 2; Length 676;
 Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;

OY 1 CPNOLSPNLMWDASTCERECSDODCAAEKCCINVCGLHSC 43
 DB 129 CP---APEKASGFPAACVESCSEADSESGVKKCCSGCG-HTC 167

RESULT 3
 507807
 MDN1 protein - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)
 C:Date: 02-Dec-1993 #sequence
 R:Dear, T.N.; Ramsay, I.A.; Kefford, R.F.
 A:Title: Differential expression of a novel gene, MDN1, in nonmetastatic rat mammary ad
 A:Reference number: 507807; MUID:88310901
 A:Accession: S07807
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Cross-references: EMBL:X13309
 R:Dear, T.N.; Kefford, R.F.
 A:Title: The MDN1 gene product is a novel member of the 'four-disulphide core' family C
 A:Reference number: JH0390; MUID:91207400
 A:Accession: JH0390
 A:Residues: 1-68 <DE2>
 A:Molecule type: mRNA
 A:Cross-references: EMBL:X13309; MUID:957500; PTDN:CAA31688.1; PTD:957501
 A:Experimental source: thymus
 C:Superfamily: 12/1
 F:23-67/Domain: antileukoproteinase repeat homology

Query Match
 Best Local Similarity 32.4%; Score 82; DB 2; Length 68;
 Matches 17; Conservative 3; Mismatches 17; Indels 6; Gaps 2;

OY 1 CPNOLSPNLMWDASTCERECSDODCAAEKCCINVCGLHSC 43
 DB 27 CPKNPPRST-----GTCVELCSGDSQCPNIQKCCSGCG-HVC 63

RESULT 4
 A47222
 Kallmann syndrome protein homolog KAL - quail

C:Species: Phasianidae gen. sp. (quail)
 C:Date: 24-Feb-1994 #sequence
 R:Legouis, R.; Cohen-Salmon, M.; del Castillo, I.; LeVilliers, J.; Capy, L.; Mornow, J.H
 A:Title: A gene deleted in Kallmann's syndrome shares homology with neural cell adhe
 A:Reference number: 517982; MUID:92018217
 A:Accession: S17982
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-680 <FRA>
 A:Cross-references: EMBL:X60299; MUID:934024; PTDN:CAA42641.1; PTD 934025
 A:Gene: GDB:KALI; KAL

A:Title: Characterization of the chicken and quail homologues of the human gene rc
 A:Accession number: A47222; MUID:94010957
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-674 <DE5>
 A:Cross-references: GB:113976; MUID:9196807; PTD:9196808
 A:Note: sequence extracted from NCBI backbone (NCBI:137995)
 F:124-170/Domain: antileukoproteinase repeat homology

Query Match
 Best Local Similarity 32.4%; Score 82; DB 2; Length 674;
 Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;

OY 1 CPNOLSPNLMWDASTCERECSDODCAAEKCCINVCGLHSC 43
 DB 128 CP---APEKASGFPAACVESCSEADSESGVKKCCSGCG-HTC 166

RESULT 5
 A40351
 adhesion-type protein ADMLX - human

C:Species: Homo sapiens (man)
 C:Date: 24-Jan-1992 #sequence
 R:Legouis, R.; Harel, J.P.; LeVilliers, J.; Clavierie, J.M.; Compain, S.; Wunderle,
 A:Title: The candidate gene for the X-linked Kallmann syndrome encodes a protein rela
 A:Reference number: A40351; MUID:92005720
 A:Accession: A40351
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-679 <LECG>
 A:Cross-references: GB:S60085; MUID:9237595; PTDN:AAB20108.1; PTD:9237597
 A:Gene: GDB:KALI; KAL
 A:Superfamily: XP24.32-XP22.32
 F:125-175/Domain: antileukoproteinase repeat homology

Query Match
 Best Local Similarity 32.0%; Score 81; DB 2; Length 679;
 Matches 15; Conservative 6; Mismatches 18; Indels 4; Gaps 2;

OY 1 CPNOLSPNLMWDASTCERECSDODCAAEKCCINVCGLHSC 43
 DB 133 CP---APEKASGFPAACVESCSEADSESGVKKCCSGCG-HTC 171

RESULT 6
 S17982
 Kallmann syndrome protein KALIG-1 - human

C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence
 R:Pratico, B.; Guicoli, S.; Pragliola, A.; Incerti, B.; Tonlorenzi, R.; Ca
 A:Title: A gene deleted in Kallmann's syndrome shares homology with neural cell adhe
 A:Reference number: 517982; MUID:92018217
 A:Accession: S17982
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-680 <FRA>
 A:Cross-references: EMBL:X60299; MUID:934024; PTDN:CAA42641.1; PTD 934025
 A:Gene: GDB:KALI; KAL

A:Cross-references: GDB:120116; OMIM:308700
 A:Map position: Xp22.32-Xp22.32
 C:Superfamily: antileukoprotease repeat homology
 E:130-176/Domain: antileukoprotease repeat homology <ALP>

Query Match	32.0%;	Score 81;	DB 2;	Length 660;
Best Local Similarity	34.9%;	Pred. No. 0.15;		
Matches	15;	Conservative	6;	Mismatches 18;
				Indels 4;
				Gaps 2

QY 1 CPNQLSPNLWDAOSTCERECRRDDCAAAEKCCINVCGLHSC 43
|| : | | : : || | | | : |
Db 134 CP---APEKASGFAACVESCEVDNECSGVKKCCSNGCG-HTC 172

RESULT 7
S61553
probable proteinase inhibitor wdml precursor - mouse
N:Alternate names: WDML protein
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
C:Accession: S61553; I49390
R:Morrison, B.W.; Leder, P.
Oncogene 9, 3417-3426, 1994
A:Title: neu and ras initiate murine mammary tumors that share genetic markers generally
A:Reference number: I48271; MUID:95060797
A:Accession: S61553
A:Molecule type: DNA; mRNA
A:Residues: 1-74 <MOR>
A:Cross-references: EMBL:X93037; NID:g1085069; PIDN:CAA63605.1; PID:g1085070
C:Genetics:
A:Gene: wdml
C:Superfamily: antileukoproteinase repeat homology
C:Keywords: extracellular protein; proteinase inhibitor
E:1-22/Domin: signal sequence #status predicted <SIG>
E:23-74/Product: probable proteinase inhibitor wdml #status predicted <MAT>
E:29-73/Domin: antileukoproteinase repeat homology <ALP>

Query Match	31.88;	Score 80.5;	DB 2;	Length 74;
Best Local Similarity	46.48;	Pred. No. 0.036;		
Matches 13; Conservative	4;	Mismatches 10;	Indels 1;	Gaps 1

```

QY      16 TCERECSRDDCAAAEKCCTINVCGLHSC 43
      ||: |: |: ||| |: |: |: |
Db      43 TCDERCTGDSGSCSGNMKCCSNGCG-HAC 69

```

RESULT 8
T25924
hypothetical protein T27E4.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 15-Oct-1999
C:Accession: T25924
R:Bradsnav, H.
Submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid T27E4.
A:Reference number: Z20111
A:Accession: T25924
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-131 <BAA>
A:Cross-references: EMBL:U64837; PIDD:AA804836.1; GSPDB:GN00023; CESP:T27E4.5
A:Experimental source: strain Bristol NZ; clone T27E4
C:Genetic8
A:Gene: CESP:T27E4.5
A:Map position: 5
A:Introns: 40/1; 53/2; 98/1; 118/3

Query Match	31.48;	Score 79.5;	DB 2;	Length 131;
Best Local Similarity	35.48;	Pred. No. 0.068;		
Matches 17;	Conservative 6;	Mismatches 20;	Indels 5;	Gaps 2

QY 1 CPNQLSPLNLWVDAQSTCE---RECRDQDC--AAAEKCCINVCGLHSC 43
||:| : ||| : :||| : :|
Db 57 CPSQLMAYFCQKISTARKTVKPCKTDQDCMFNVQKCCDAGCGFNV 104

```

RESULT      9
TITROR
basic proteinase inhibitor - loggerhead
C:Species: Caretta caretta (loggerhead)
C:Date: 31-May-1979 #sequence_revision 23-Oct-1981 #text_change 05-Aug-1994
C:Accession: A01224
R:Kato, I.: Tomienna, N.
Fed. Proc. 38, 832, 1979
A:Title: Trypsin--subtilisin inhibitor from red sea turtle eggwhite consists of two t
A:Reference number: A01224
A:Accession: A01224
A:Molecule type: protein
A:Residues: 1-110 <RKF>
C:Comment: This inhibitor, isolated from egg white, consists of two nonhomologous dom
C:Superfamily: loggerhead basic proteinase inhibitor; animal kunitz-type proteinase i
C:Keywords: pyroglutamic acid; serine proteinase inhibitor
F:8-58/Domain: animal kunitz-type proteinase inhibitor homology <BPI>
F:63-105/Domain: antileukoproteinase repeat homology <ALP>
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:8-58,17-41,33-54,67-93,76-97,80-92,86-101/Disulfide bonds: #status predicted
F:18/Inhibitory site: Lys (trypsin) #status predicted

```

Query Match	30.6%;	Score 77.5;	DB 1;	Length 110;
Best Local Similarity	33.3%;	Pred. No. 0.099;		
Matches 13;	Conservative 3;	Mismatches 16;	Indels 7;	Gaps 1

```

QY      1 CPNDLSPNLWDAQSTICREKSCSDQDCAAEKCCINWCG 39
          || : | | | | : || : ||
DB      67 CPKTSGPGI-----CLHGCDSDSDCKEKGQCKCPDGGC 98

```

RESULT 10
154768
epididymis-specific four-disulfide core protein CE4 - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 24-Oct-2000
C:Accession: 154768
R:Ellerbrock, K.; Pera, I.; Hartung, S.; Ivell, R.
Int. J. Androl. 17, 314-323, 1994
A:Title: Gene expression in the dog epididymis: a model for human epididymal function
A:Reference number: 154768; MUID:95265175
A:Accession: 154768
A:Status: preliminary;
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-124 <ELL>
A:Cross-references: GB:S77395; NID:g945180; PIDN:ABB34264.1; PID:g945181
A:Superfamily: antileukoproteinas; antileukoproteinas repeat homology
;76-123/Domain: antileukoproteinas repeat homology <ALP>

Query Match	29.8%	Score 75.5;	DB 2;	Length 124;
Best Local Similarity	37.8%	Pred. No. 0.18;		
Matches 17;	Conservative 2;	Mismatches 19;	Indels 7;	Gaps 2

OY 1 CP--NQLSPNLWDAQSTCERECRSRDQDCAAAEKCCINVCGLHSC 43
||| | : : ||| ||| ||
80 CPQVNTDFPQL-----GLCDQDCQVDSHCPGLLKCCYNCGCKVSC 119

Db

```

RESULT 11
S25454
epididymis-specific four-disulfide core protein HE4 - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Oct-2000
C:Accession: S25454
C:Kirchoff, C.; Haben, I.; Ivell, R.; Krull, N.

```

Biol. Reprod. 45, 350-357, 1991
 A:Title: A major human epididymis-specific cDNA encodes a protein with sequence homology
 A:Reference number: S25454; MUID:92155963
 A:Accession: S25454
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-125 <KTR>
 A:Cross-references: EMBL:X63187; NID:932050; PID:CA44869.1; PID:g32051
 C:Genetics:
 A:Gene: GDB:HE4
 A:Cross-references: GDB:9956898
 A:Map position: 20q12-20q13.2
 C:Superfamily: antileukoproteinasase; antileukoproteinasase repeat; homology <ALP1>
 F:32-74/Domain: antileukoproteinasase repeat; homology <ALP2>
 F:77-123/Domain: antileukoproteinasase repeat; homology <ALP2>

	Query Match	29.4%	Score 74.5	DB 2	length 125;
	Best Local Similarity	33.3%	Pred. No. 0.23;		
	Matches	16; Conservative	2; Mismatches	25; Indels	5; Gaps
QY	1 CPNQL-----SPNLAVDAQSTCERCSCSRDQCAAAEKCCINVCGLHSC	43			
	:: :: :: ::				
Db	73 CPNDKGSGCPVYNINFPLGICRDGCQVDYDQCPGMKKCCRNCGGVSC	120			

RESULT 12
A24178
Whey acidic protein - Arabian camel
C:Species: Camelus dromedarius (Arabian camel)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Aug-1997
C:Accession: A24178
R:Beig, O.U.; von Bahr-Lindstrom, H.; Zaidi, Z.H.; Jorjwall, H.
Eur. J. Biochem. 159, 195-201, 1986
A:Title: A camel milk whey protein rich in half-cystine. Primary structure, assessment
A:Reference number: A24178; MUID:86300719
A:Accession: A24178
A:Molecule type: protein
A:Residues: 1-117 <BEG>
C:Superfamily: antileukoproteinasase: antileukoproteinasase repeat homology
F:9-54/Domain: antileukoproteinasase repeat homology <ALP2>
F:64-111/Domain: antileukoproteinasase repeat homology <ALP1>

Query Match	29.2%	Score 74	DB 2	Length 117
Best Local Similarity	31.1%	Pred. No. 0.24		
Matches 14	Conservative 5	Mismatches 14	Indels 12	Gaps 2
QY	1 CPNDLSRWLWDAASTC-----ERECSSDDQDCAAEKCCINVCGL 40			
Db	68 CP-----WQPTPLTAHCKLEKNDKSSDDQCEGKHKCCFFSSCAM 105			

```

RESULT 13
JEO252
trappin-6 - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 16-Jul-1999
C:Accession: JEO252
R:Puritant, Y.; Kato, A.; Yasue, H.; Alexander, L.J.; Beattie, C.W.; Hirose, S.
J. Biochem. 124, 491-502, 1998
A:Title: Evolution of the trappin multigene family in the sulidae.
A:Reference number: JEO250; MUID:98391820
A:Accession: JEO252
A:Molecule type: mRNA
A:Residues: 1-96 <FUR>
A:Cross-references: DDBJ:AB003283
C:Superfamily: antileukoproteinasase repeat homology
F;51-96/Domain: antileukoproteinasase repeat homology <ALP>

```

Query Match	28.98;	Score 73;	DB 2;	Length 96;
Best Local Similarity	37.58;	Pred. No. 0.27;		

	Matches	15;	Conservative	4;	Mismatches	15;	Indels	6;	Gaps	2;
QY	1	CPN-QLSRPLWYDAOSTCERECSRQDDCAAAKCCINVCG	39							
Db	55	CPRVVHICNLW-----NPNCQMDRAHCPCGAKKCCGEGFCG	89							

RESULT 14
B35752

caltrin-like protein II - guinea pig

C:\Date: 05-Oct-1990 #sequence revision
C:\Species: Cavia porcellus (guinea pig)

```
C;Accession: B35752
```

R;Coronel, C.E.; Sa

J. Biol. Chem. 265, 6854-6859, 1990

A; Reference number: A35752; MUID:90216715

A;Accession: B35752

A;Status: preliminary
A;Molecule type: protein

```

n/molecule type: protein
A;Residues: 1-55 <COR>

```

C: Superfamily: antileukoproteinase repeat homology

F;10-55/Domain: antileukoprotease repeat homology <ALP>

Query Match	28.3%	Score 71.5;	DB 2;	Length 55;
Best Local Similarity	33.3%	Pred. No. 0.26;		
Matches 13;	Conservative 4;	Mismatches 11;	Indels 11;	Gaps 1;

```

QY      1 CPNDLSPNLWDAQSTICERECGRDODCAAAEKCCTNVCG 39
        || : || : || : || : || : || : || : ||
Db      21 CPANHPRN-----KTSYDYDCKRPQKCPGYCG 48

```

RESULT 15
S47040
gene Tt52 protein - *Trichuris trichiura* (fragment)
C:Species: *Trichuris trichiura*

```
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 16-Jul-1999
C;Accession: E47040
```

C/Accession: 347040
R:Drake, L.: Brooks, H.: Barker, G.C.: Bundy, D.A.P.

submitted to the EMBL Data Library, April 1994

A;Reference number: S47040

A;Accession: 547040
A;molecule type: mRNA

A;Residues: 1-478 <DRA>

A;Cross-references: EMBL:X78981; NID:g520892; PID:g520893

E:131-174/Domain: anti-leukoprotease repeat homology
c;superfamily: anti-leukoprotease repeat homology

F;232-275/Domain: antileukoproteinnase repeat homology <ALP>

Query Match	27.1%;	Score 68.5;	DB 2;	Length 478;
Best Local Similarity	47.8%;	Pred. NO. 2.6;		
Matches 11;	Conservative 2;	Mismatches 9;	Indels 1;	Gaps 1;

QY	21	CSRQDCCAAAEKCCINVCGLHSC	43
		:	
Db	98	CNSDTECKGTEKCCLTAG-HQC	119

```
Search completed: February 26, 2002, 01:29:33
Job time: 412 sec
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:41 ; Search time 162.6 Seconds
(without alignments)
54.211 Million cell updates/sec

Title: US-09-819-136-2_COPY_39_157

Perfect score: 692

Sequence: 1 PNLWDASTGECRSRPDQ.....RCYDAEACGLHLHYPC 119

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_1101:*

- 1: /SID2/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	629	90.9	216	22	Human membrane or
2	125.5	18.1	319	16	Follistatin. Syn
3	117.5	17.0	195	21	Balanus amphitrite
4	116.5	16.8	304	13	Human EDF-binding
5	116.5	16.8	317	10	Human folliostat
6	116.5	16.8	317	22	Human folliostat
7	116.5	16.8	344	22	Human polypeptide
8	116.5	16.8	345	22	Human folliostat
9	116.5	16.8	432	22	Human polypeptide
10	115.5	16.7	308	18	Follistatin relate
11	113.5	16.4	304	13	Plg EDF-binding pr

12	113.5	16.4	304	13	AA20062	Rat EDF-binding pr
13	111.5	16.1	233	20	AA14569	Rat apoptosis inh
14	111.5	16.1	226	20	AA14570	Rat apoptosis inh
15	104.5	15.1	815	21	AA19728	Human SECX Clone A
16	104.5	15.1	842	21	AA19727	Human SECX Clone 4
17	104.5	15.1	842	22	AA182300	Human folliostat
18	104.5	15.0	413	22	AA193754	Human protein seg
19	101.5	14.7	263	20	AA14571	Human apoptosis in
20	101.5	14.7	263	20	AA106201	Follistatin module
21	101.5	14.7	263	20	AA101098	Human folliostat
22	101.5	14.7	263	22	AA1012413	Human PRO1308 poly
23	101.5	14.7	263	22	AA172875	Human PRO1308 prot
24	101.5	14.7	263	22	AA172875	Human folliostat
25	101.5	14.7	263	22	AA131190	Amino acid sequenc
26	101.5	14.7	263	22	AA150987	Human PRO1308 prot
27	99.5	14.4	1508	20	AA127141	Human slit-1 matur
28	99.5	14.4	1508	20	AA104138	Human slit-1 matur
29	99.5	14.4	1508	20	AA104138	Human slit-1 matur
30	99.5	14.4	1534	19	AA165966	Amino acid sequenc
31	99.5	14.4	1534	20	AA127144	Human slit-1 prote
32	99.5	14.4	1534	20	AA104139	Human slit-1 prote
33	99.5	14.4	1534	20	AA106707	Human sequence o
34	99.5	14.3	71	20	AA106202	Human FCMF follist
35	97.5	14.1	1496	20	AA127143	Human slit-3 matur
36	97.5	14.1	1496	20	AA104136	Human slit-3 matur
37	97.5	14.1	1523	20	AA127146	Human slit-3 prote
38	97.5	14.1	1523	20	AA141412	Human slit-3 prote
39	97.5	14.1	1523	20	AA104137	Human slit-3 prote
40	97.5	14.1	1523	21	AA199395	Human PRO1336 (Inv
41	97.5	14.1	1523	22	AA166144	Protein of the Inv
42	96	13.9	499	21	AA107738	A snake venom prot
43	96	13.9	781	21	AA107738	A snake venom prot
44	96	13.9	781	21	AA107705	Amino acid sequenc
45	96	13.9	790	21	AA107742	A snake venom prot

ALIGNMENTS

RESULT	1
AA188333	standard; Protein: 216 AA.
ID	AA188333
AC	AA188333
DT	23-MAY-2001 (first entry)
DE	Human membrane or secretory protein clone PSEC0040.
XX	Human, secretory protein; membrane protein; vaccine; gene therapy;
KW	rheumatoid arthritis; diabetes.
XX	
OS	Homo sapiens.
PN	EP1067182-A2.
XX	
PD	10-JAN-2001.
XX	
PF	07-JUL-2000; 2000EP-0114090.
XX	
PR	08-JUL-1999; 99JP-0194179.
PR	11-JAN-2000; 2000JP-0118775.
XX	
PR	02-MAY-2000; 2000JP-0183766.
XX	
PI	(HELI-) HELIX RES INST.
XX	
DR	Oca T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX	WPI; 2001-093989/11.
DR	N-PSDB; AAF93760.
XX	
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -

XX Claim 1; SEQ ID 34; 609pp + CD ROM; English.

PS This invention relates to nucleic acid sequences AAF93744 - AAF93916

XX which encode human secretory or membrane proteins represented by

CC AAB88317 - AAB88419. Included in the invention are primers

CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the

CC cDNA sequences of the invention. The invention also includes methods for

CC the production of antibodies directed against the proteins, and cDNA

CC sequences, which can be used in vaccines. The polynucleotide sequences

CC can be used in gene therapy. The polynucleotide sequences and the

CC proteins they encode may be used in the prevention, treatment and

CC diagnosis of diseases associated with inappropriate secretory

CC protein/membrane protein expression. The nucleic acids and complementary

CC sequences may also be used as DNA probes in diagnostic assays

CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the

CC presence of similar nucleic acid sequences in samples. They may also be

CC used to study the expression and function of secretory proteins/membrane

CC polypeptides and their role in metabolism. The polypeptides may be used

CC as antigens in the production of antibodies against them and in assays to

CC identify modulators (agonists and antagonists) of expression and

CC activity. The antibodies and antagonists may also be used as therapeutic

CC agents to down regulate expression and activity. The antibodies may also

CC be used as diagnostic agents for detecting the presence of the

CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay

CC (ELISA). Examples of diseases which may be treated include rheumatoid

CC arthritis and diabetes.

XX

Sequence 216 AA;

Query Match 90.9%; Score 629; DB 22; Length 216;

Best Local Similarity 100.0%; Pred. No. 2.4e-46;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PULWVDAOSTCERECSDODCAAEKCCINVGCLHSCVAARFPSPAPATTAASCEGFC 60

DB 39 PULWVDAOSTCERECSDODCAAEKCCINVGCLHSCVAARFPSPAPATTAASCEGFC 98

OY 61 POGGSDCDIMDGOVPCRCRCEKPEPFTCASDGLTYVRCYMDAEAC 108

DB 99 POGGSDCDIMDGOVPCRCRCEKPEPFTCASDGLTYVRCYMDAEAC 146

RESULT 2

AAR74600 standard; Protein; 319 AA.

AC AAR74600;

XX 24-OCT-1995 (first entry)

DE Follistatin.

XX Neuronal differentiation; transforming growth factor; TGF-beta;

XX degenerative disorders; nervous system; Alzheimer's disease;

KW Parkinson's disease; amyotrophic lateral disease; Pick's disease;

KW Huntington's disease; multiple sclerosis; anoxia-ischaemia; trauma;

KW ageing; tachycardia; atrial cardiac arrhythmia.

XX Synthetic.

OS

XX W09510611-A.

PN 20-APR-1995.

XX 14-OCT-1995; 95WO-1107450.

XX 14-OCT-1993; 93US-0136748.

XX (HARD) HARVARD COLLEGE.

PA Melton D, Hemmati-Briyaniou A;

PI

XX WPI: 1995-178527/23.

DR N-PSDB; AAQ90017.

XX

PT Inducing neuronal differentiation using TGF-beta family agonist -

PT useful to prevent death and/or degeneration of neuronal cells, e.g.

PT to treat Alzheimer's disease, multiple sclerosis, etc.

XX Example; Page 41; 53pp; English.

XX

XX The sequence is that of follistatin. This protein can be used to

CC differentiate neuronal cell phenotype by antagonising the action of

CC at least one member of the transforming growth factor beta (TGF-beta)

CC family which normally induces the cell to differentiate to a non-

CC neuronal phenotype. The agent can be used to prevent death and/or

CC degeneration of a neuronal cell in a degenerative disorder of the

CC nervous system, e.g. Alzheimer's disease, Parkinson's disease,

CC amyotrophic lateral sclerosis, Pick's disease, Huntington's disease,

CC multiple sclerosis, neuronal damage resulting from anoxia-ischaemia

CC or trauma and neuronal degeneration associated with ageing.

XX

Sequence 319 AA;

Query Match 18.1%; Score 125.5; DB 16; Length 319;

Best Local Similarity 26.2%; Pred. No. 0.0023;

Matches 42; Conservative 11; Mismatches 60; Indels 47; Gaps 8;

OY 1 PULWVDAOSTCERECSDODCAAEKCCINVGCLHSCVAARFPSPAPATTAASCEGFC 49

DB 154 PELDVYGYGKCKKTCRDVLCPSSSCVVDGYNAYCVLCNRCIPESPDPYLGNDGI 212

OY 50 -----TTAASCEGFCVPCOGSDCDIMD-----GQVYCR-C 78

DB 213 TYGSACHLRKATCGLIGSLAYEGKCKAKSCEDIQC-SAGKXC-IWDSRVYGRCALC 270

OY 79 RDRC--EKPEPFTCASDGLTYVRCYMDAEACLRGLHFI 116

DB 271 dLlGeskSdtdvcsadnttYpsecamkqacstglllev 310

RESULT 3

AAB23270 standard; Protein; 195 AA.

AC AAB23270;

XX 02-FEB-2001 (first entry)

DE Balanus amphitrite adhesion/metamorphosis-related protein Bcs-4.

XX Adhesion/metamorphosis-related protein Bcs-4; barnacle; larva-specific;

KW adhesion inhibition; metamorphosis inhibition; compound screening;

KW antifouling composition.

XX Balanus amphitrite.

OS JP2000226985-A.

PN 22-AUG-2000.

XX 09-FEB-1999; 99JP-0031067.

XX 09-FEB-1999; 99JP-0031067.

XX (KAITO) KAITO BIOTECHNOLOGY KENKYUSHO KK.

PA WPI: 2000-649634/63.

DR N-PSDB; AAA97604.

XX Novel barnacle klpis larva-specific adhesion/metamorphosis-related

PT gene useful for identifying inhibitors of barnacle adhesion -

XX

Query Match	16.8%	Score 116.5	DB 13	Length 304
Best Local Similarity	25.9%	Pred. No. 0.013		
Matches 42	Conservative 11	Mismatches 58	Indels 51	Gaps 8

CC A testicular lambda gt11 cDNA library was screened with a cDNA probe
CC encoding the first 317 amino acids of the porcine FS (pFS) precursor.
CC 12 positive clones were obtained. From the nucleotide sequences of
CC the clones, it was determined that there is encoded an hFS precursor
CC containing a 344 amino acid sequence which is highly homologous with

CC the 344-residue pFS precursor. It differs from the 344-residue pFS
 CC precursor by only 6 residues (see FT). AAP93396 is encoded by bases
 CC 28-978 of AAN90641.
 XX
 SO Sequence 317 AA:

Query Match 16.8%; Score 116.5; DB 10; Length 317;
 Best Local Similarity 25.9%; Pred. No. 0.013;
 Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8;

QY 1 PNLMWAQSTCEKESRDQDCAAEKCCINWGLHSCVAA-RFPSPAP----- 49
 DB 154 pelyvgggrcktc-rdvfcpgstcvvdqtnaycvtcnrlcpepassseqylegndgv 212
 OY 50 -----TTASCEGFVCPQGSDDCIDMD---GQPYCR-- 77
 DB 213 tyssachlrkacdlgrslglaegkclakscedlqc-tgqkcc-lwdfkvgirgrcslc 270
 QY 78 ---CRDCKEKEPSFTCASDGLTYNRCYMDAECRLGLHLHI 116
 DB 271 delcpdsksdcp--vcasdnatyasacamkaeacssgyllev 310

RESULT 6
 AAY72630
 ID AAY72630 standard; Protein; 317 AA.

AC AAY72630;
 XX
 DF 02-MAY-2001 (first entry)
 XX

DE Human follistatin-1 protein.

KM Human; follistatin; cytostatic; vulnery; arthritis; immunomodulatory;
 KM therapy; liver cirrhosis; Gaucher's disease; vaccine; gene therapy;
 KM reproductive system disorder; cell growth disorder; Digorge syndrome;
 KM tumour; inflammatory bowel disease; anaemia; autoimmune disorder; AIDS;
 KM Acquired Immune Deficiency Syndrome; SLE; systemic lupus erythematosus;
 KM Wiskott-Aldrich disorder; cardiovascular disorder; cancer; ischaemia;
 KM Kaposi sarcoma; hyperproliferative disorder; fibrotic disorder; wound;
 KM neurodegenerative disorder; congestive heart failure; Crohn's disease;
 KM Alzheimer's disease; Parkinson's disease; pulmonary fibrosis.
 XX

OS Homo sapiens.

XX WO200105998-A1.

XX 25-JAN-2001.

XX 14-JUL-2000; 2000WO-US19198.

XX 16-JUL-1999; 99US-0144088.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Duan RD, Ruben SM;

XX WPI; 2001-103150/11.

XX Nucleic acids encoding follistatin-3, useful for the prevention,
 PT diagnosis and treatment of e.g. arthritis, liver cirrhosis and
 PT Gaucher's disease -
 XX

XX Disclosure; Page 232-233; 245pp; English.

XX The present sequence is human follistatin-1 protein, a member of
 CC inhibin-related proteins. Follistatin-1 protein is 43.2% identical
 CC to follistatin-3 protein and is an important factor in the regulation
 CC of follicle development and spermatogenesis in reproductive systems.
 CC It also acts as an antagonist of activin by preventing the interaction
 CC of activin with its receptor. In a similar manner follistatin-1
 CC also targets TGF-beta superfamily members. Follistatin-3 sequences are

CC used in the prevention, diagnosis and treatment of diseases associated
 CC with inappropriate expression of follistatin-3. The follistatin-3
 CC sequences and their agonists or antagonists are useful in the diagnosis,
 CC prevention and treatment of reproductive system-related disorders; cell
 CC growth and differentiation disorders (tumours, arthritis, inflammatory
 CC bowel disease); immune disorders (Digorge syndrome, anaemia and
 CC Wiskott-Aldrich disorder); autoimmune disorders (AIDS, Crohn's disease,
 CC systemic lupus erythematosus-SLE); hyperproliferative disorders (pustula,
 CC Gaucher's disease, Sezary syndrome); cardiovascular disorders
 CC (pericarditis, congestive heart failure, ischaemia); cancers (Kaposi
 CC sarcoma, lymphoma); neurodegenerative disorders (Alzheimer's disease,
 CC Parkinson's disease); to stimulate wound healing and treat other fibrotic
 CC disorders (liver cirrhosis and pulmonary fibrosis) and inhibit
 CC angiogenesis. Follistatin-3 DNA sequence is also useful in chromosome
 CC identification and in gene therapy.
 XX

SO Sequence 317 AA;

Query Match 16.8%; Score 116.5; DB 22; Length 317;
 Best Local Similarity 25.9%; Pred. No. 0.013;
 Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8;

QY 1 PNLMWAQSTCEKESRDQDCAAEKCCINWGLHSCVAA-RFPSPAP----- 49
 DB 154 pelyvgggrcktc-rdvfcpgstcvvdqtnaycvtcnrlcpepassseqylegndgv 212
 QY 50 -----TTASCEGFVCPQGSDDCIDMD---GQPYCR-- 77
 DB 213 tyssachlrkacdlgrslglaegkclakscedlqc-tgqkcc-lwdfkvgirgrcslc 270
 QY 78 ---CRDCKEKEPSFTCASDGLTYNRCYMDAECRLGLHLHI 116
 DB 271 delcpdsksdcp--vcasdnatyasacamkaeacssgyllev 310

RESULT 7
 AAM39389
 ID AAM39389 standard; Protein; 344 AA.

XX AAM39389;

XX 22-OCT-2001 (first entry)
 DT

DE Human polypeptide seq ID NO 2534.

KM Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia.
 XX

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	WPI: 2001-442253/47.
DR	N-PSDB; AAI58545.
XX	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
PS	Example 4; SEQ ID NO 2534; 10078bp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM8642-AA42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	centralised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilization of the activities such as: Immune system suppression,
CC	Activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SO	Sequence 344 AA;
Query Match	16.8%; Score 116.5; DB 22; Length 344;
Best Local Similarity	25.9%; Pred. No. 0.014;
Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8	
QY	1 PRLWDAOSTGERESRRODDCAAEKCCINNCGHSCVAA-RFPGSPAP----- 49
DB	154 PELVYGYGRCKKCRDVFQPSATCVDQIMNAYCVFCMTCIPCPASSQYLGNQDV 212
QY	50 -----TFAASCEGFVCPQGSDDCDIMD---GQPYCR-- 77
DB	213 TYSSECHLRKATCIIIGRIIGLAGEYKCLKAKSCEDIGC-TGGRKCR-LWDFKVGIRGCSIC 270
QY	78 ---CRDRCEKEPSFTCASDGLTYNYKRCYMDAEACILGHLHI 116
DB	271 DELCPDASKSDP--VCASDNLTYASECMKEAKCASQVILLEV 310
RESULT 8	
AAID	AAI70351
XX	AAI70351 standard; protein; 345 AA.
AC	AAI70351;
XX	
DT	30-APR-2001 (first entry)
XX	
DE	Human follistatin FS288 amino acid sequence SEQ ID NO:1.
XX	
KW	Human: follistatin; FS288; follistatin antagonist; activin.
OS	Homo sapiens.
XX	
FH	Key
FT	Peptide
FT	Domain
FT	Domain
FT	Misc-difference
FT	Domain
FT	Label- I
FT	Label- heparin-sulphate-site
FT	Label- II

FT	Domain	242..318
FT	/label= III	
FT	Domain	319..345
FT	/label= C-domain	
PN	M0200109368-A1.	
PD		
XX	08-FEB-2001.	
XX		
PF	27-JUL-2000; 2000WO-US20444.	
XX		
PR	30-JUL-1999; 99US-0146576.	
XX		
PA	(GEHO) GEN HOSPITAL CORP.	
PI	Kautmann HT, Schneyer A, Sidsis Y, Sluss P;	
DR	WPI: 2001-182970/1B.	
XX		
PT	Novel follistatin antagonist having non-activin binding follistatin	
PT	N-domain or non-activin binding FLRGN domain, in which tryptophan	
XX	residue at position 4 or 36 of the domains is substituted by alanine -	
PS	Claim 1; Fig 1; 34pp: English.	
XX		
CC	The present invention describes a follistatin antagonist (I) comprising	
CC	residues 3-63 (FI) of a 315 amino acid non-activin binding (NAB)	
CC	follistatin N-dominant (Ia) sequence (see AAB70351) or residues 3-61 (R2)	
CC	of a 61 amino acid NAB FLRG domain (Ib) sequence (see AAB70352), and a	
CC	heparin sulfate-binding moiety (IC). The tryptophan residue at position	
CC	4 or 36 of (Ia) or (Ib) can be substituted by an amino acid of A, G, L,	
CC	I, V, P, S, T, M, C, N, Q, D, E, R, Y and H. Also described are: (1) a	
CC	nucleic acid (II) encoding (I); (2) a vector (III) comprising (II); and	
CC	(3) a host cell (IV) transformed with (III). (I) is useful for increasing	
CC	activin activity in an activin-sensitive tissue e.g. a prostate tissue, on	
CC	contact with cells of the tissue. The contact process is preferably	
CC	carried out by expressing (I) in vivo from a vector comprising (II).	
CC	The FS antagonist displays a structure extremely similar to natural,	
CC	native FS. Because of the similarity, the cellular machinery treats the	
CC	FS antagonist as if it were native FS, a molecule naturally present in	
CC	the cell. Consequently, the FS antagonist causes minimal disruption to	
CC	normal cellular processes, i.e. the FS antagonist's side effects are	
CC	minimised. In addition, because the cell treats the FS antagonist as if	
CC	it were native FS, the antagonist undergoes any FS-specific transport or	
CC	localisation that takes place in the target cell or target tissue. Thus	
CC	the FS antagonist is similarly localised to maximise its competition	
CC	with the native FS.	
XX		
SQ	Sequence 345 AA;	
	Query Match 16.8%; Score 116.5; DB 22; Length 345;	
	Best Local Similarity 25.9%; Pred. No. 0.014; Indels 51; Gaps 8	
	Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps	
QY	1 PNLMWAOSTERECSRPDCAAAAEKCCINNGLSCVAA-BFPGSPAP----- 49	
	: : : :	
Db	155 pelervgvgrcktc-rdvfcpgstcvcvdgmnyrcvtncilpepaseqylgndgv 213	
	: : : :	
QY	50 -----TTASCEGFVCPOGSGDCDIWD--GQPYCR-- 77	
	: : : :	
Db	214 tyssachlrkatcllgrislglayegkcklkakecdlqc-tggkkc-lwdfkvrgfscslc 271	
	: : : :	
QY	78 ---CDRCCKEPEFPCASDGSLTYRYRCVADACRLGLLHI 116	
	: : : :	
Db	272 delcpdsksdcp--vcasdnatlyasecamkeaacsgylllev 311	
	: : : :	
RESULT	9	
ID	AAM41175 standard; Protein: 432 AA.	
XX		
CC	AAM41175:	

XX	22-OCT-2001	(first entry)
DT		
XX		
DE		
XX		
XX	Human polypeptide SEQ ID NO 6106.	
KW	Human; neurotropic; immunosuppressant; cytotactic; gene therapy; cancer;	
KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;	
KW	leukemia.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200153312-A1.	
XX		
PD		
XX	26-JUL-2001.	
XX		
PF		
XX	26-DEC-2000; 2000WO-US34263.	
XX		
XX	21-JAN-2000; 2000US-0488725.	
PR	25-APR-2000; 2000US-0552317.	
PR	09-JUL-2000; 2000US-0598042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
XX	29-NOV-2000; 2000US-0727344.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
DR	WPI: 2001-442253/47.	
XX		
DR	N-PSDB; AA160331.	
XX		
PT	NOVEL nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
XX		
PS	Example 2; SEQ ID NO 6106; 10078bp; English.	
XX		
XX		
CC	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	the encoded polypeptides (AA836642-AA442213) with neurotropic,	
CC	immunosuppressant and cytotactic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukemias and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
XX		
XX	Sequence 432 AA:	

```

Query Match          16.8%; Score 116.5; DB 22; Length 432;
Best Local Similarity 25.9%; Pred.No.0.018;
Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8;

QY 1 PNLWDAOSTERCERSRQDCAAEKCCINWGLHSCVAA--RFPSPAP-----49
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 242 pelvqyagrcktc-rdvtcpbstcvvdqnnaycvtcmticpceasseqylcgndiv 300
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 50 -----ITASCSGEFVCPQGSDDCDIND--GQPYCR--77
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

```

Db      301  lysaachlrlkaccllgrsigrlayegckclkskscedlge-tggkkc-lwdflvrgjrsclc 335
QY      78  ---CRDRCCKEPEFTCASGLTYTNRCYMDAEACLRLGHLHI 116
Db      359  delcpdsksdpe--vcasdnatlyasecamkeacssgylliev 398

RESULT 10
AA017860
AA017860 standard; Protein: 308 AA.
AA017860;
AA017860;
04-FEB-1998 (first entry)
Follistatin related protein..
Rheumatoid arthritis; auto-antigen; FRP; diagnosis;
prediction; synovial cell; follistatin related protein.
Homo sapiens.
Key      Location/Qualifiers
FH      Misc-difference 175
          /note="encoded by A?"
PN      WO9717441-A1.
PD      15-MAY-1997.
PF      06-NOV-1996; 96WO-JP03250.
PR      07-NOV-1995; 95JP-0288957.
PA      (KANF ) KANEKA CORP.
PI      Kishimura M, Nakao K, Osakada F, Osaki S, Tanaka M;
DR      WPI; 1997-281030/25.
PS      N-PSDB; AAT68831.
PT      Auto-antigen from synovial cells of rheumatoid arthritis patients -
PR      binds to antibodies present in these patients, for diagnosis and
PI      prediction of the disease .
PS      Claim 6; Pages 38-40; 61pp; Japanese.
XX      The present sequence is the rheumatoid arthritis (RA)
XX      auto-antigen follistatin related protein (FRP), which can be used
XX      to diagnose and predict the development of RA by reaction with
XX      antibodies in biological specimens, e.g., sera, from patients.
XX      RNA was isolated from synovial cells from a RA patient and used to
XX      construct a cDNA library. This was screened using igg separated
XX      from the synovial fluid of a RA patient. Active clones were
XX      isolated in a cloning vector, and inserted into an expression
XX      vector for the transformation of E. coli NM522. Transformsants on
XX      culture express clone A peptide and FRP into the culture medium.
XX      Sequence 308 AA;
XX

```

```

Query Match          16.7%;   Pred. 115.5; DB 18; Length 308;
Best Local Similarity 23.5%;   Pred. No. 0.016;
Matches 27; Conservative 16; Mismatches 31; Indels 41; Gaps 4

QY      4 WDAOSTGERECSRDPDAAAEKCCINV-CGLHSCVAARFPESPAPPTAASCEGVCPQ 62
      ||| :::: | || |
Db      17 wvra -----eeelrskskicanvfq----- 37

QY      63 QGSDCDIWD-GOPYRCRDRCKEEPSFCASDGLTYNNRCYMDAEACLRGLLHI 116
      | : : : | : : : | : | | | : : : | : | | : :
Db      38 agrecaavekepeptcliegckphkrpyvcsgnygkylmhceimdradcltyskiqv 92

```


PI	Ohnawa Y, Uchiyama Y;
XX	
DR	WPI: 1999-395180/33.
DR	N-PSDB: AAX79359.
XX	
PT	New protein obtained by culturing pheochromocytoma transformed by a
PT	bcl2 gene as a protooncogene, useful in remedies for
PT	apoptosis-caused diseases. e.g. Alzheimer disease
XX	
PS	Claim 3; Page 38-39; 50pp; Japanese.
XX	
CC	This sequence represents the mature, apoptosis inhibitory factor PCNF35.
CC	The coding sequence was isolated from rat pheochromocytoma PC12 cells
CC	transformed with the human bcl2 gene. The protein can be used for the
CC	treatment of apoptosis-caused diseases, including those due to nerve
CC	cell death, e.g. Parkinson's disease and Alzheimer's disease.
XX	
SO	Sequence 233 AA;

Query Match	16.1%;	Score 111.5;	DB 20;	Length 233;
Best Local Similarity	26.8%;	Pred. No. 0.026;		
Matches	34;	Conservative	15;	Mismatches 39;
			Indels	39;
			Gaps	7;

RESULT	14
AAV14570	
ID	AAV14570 standard; Protein; 256 AA.
XX	
AC	AAV14570;
XX	
DT	14-SEP-1999 (first entry)
XX	
DE	Rat apoptosis inhibitory factor PCTF35 protein.
XX	
KW	Neotrophic; neuroprotective; apoptosis inhibitory factor; PCTF35; rat;
KW	pheochromocytoma cell; PC12 cell; human; bcl12; nerve cell death;
KW	Parkinson's disease; Alzheimer's disease.
OS	Rattus sp.
XX	
PN	WO9931237-A1.
XX	
PD	24-JUN-1999.
XX	
PF	11-DEC-1998; 98WO-JP05609.
XX	
PR	14-MAY-1998; 98JP-0131634.
PR	12-DEC-1997; 97JP-0343112.
XX	
PA	(TAIS) TAISHO PHARM CO LTD.
XX	
PI	Ohsawa Y, Uchiyama Y;
XX	
DR	WPI; 1999-395180/33.
DR	N-PSDB; AAX79360.
XX	
PT	New protein obtained by culturing pheochromocytoma transformed by a
PT	bcl2 gene as a protooncogene, useful in remedies for
XX	apoptosis-related diseases. e.g. Alzheimer disease

PS Example; Page 39-43; 50pp; Japanese.
XX
CC This sequence represents the full length apoptosis inhibitory factor
CC PCPp33. The coding sequence was isolated from rat pheochromocytoma PC12
CC cells transformed with the human bcl2 gene. The protein can be used for
CC the treatment of apoptosis-caused diseases, including those due to nerve
CC cell death, e.g. Parkinson's disease and Alzheimer's disease.
XX
XX Sequence 256 AA;
50

Query Match	16.1%;	Score 111.5;	DB 20;	Length 256;
Best Local Similarity	26.8%;	Pred. No. 0.029;		
Matches 34;	Conservative 15;	Mismatches 39;	Indels 39;	Gaps 7;

	RESULT	15
AAAB19728	ID	AAAB19728 standard; Protein; 815 AA.
XX		
AC	AAAB19728;	
XX		
DT	19-FEB-2001	(first entry)
XX		
DE	Human SECK clone AC012614_1.0.0.123 encoded protein.	
XX		
KW	SECK; human; diagnosis; therapy; cell adhesion;	
KM	neurodegenerative disease; epilepsy; tissue regeneration.	
OS	Homo sapiens.	
XX		
PN	MOZ00061154-A2.	
XX		
PD	19-OCT-2000.	
PE	07-APR-2000; 200OWO-US09392.	
XX		
PR	09-APR-1999; 99US-0128514.	
XX		
PP	03-MAR-2000; 200OUS-0128514.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX		
P1	Fernandez E, Vernet C, Shinkets R;	
D1	WP1: 2000-679487/66.	
DR	N-PDB; AAA88797.	
XX		
FT	SECK polypeptides and the nucleic acids that encode them, useful for	
PT	diagnosing, preventing and treating e.g. cancers, inflammation,	
XX	arthritis and immunological disorders -	
PS	Claim 5; Fig 9; 143pp; English.	
CC	The present sequence is that of the protein encoded by novel SECK	
CC	clone AC012614_1.0.123 (see AAA88797). It is a surface adhesion	
CC	protein-like variant expressed in glioma, osteoblast, lung carcinoma	
CC	and other cancer cells, and small intestine. The invention provides	
CC	a novel SECK polynucleotides (see AAA88789-804) and the secreted or	
CC	membrane-associated proteins encoded by them (see ABA19720-34).	
CC	SECK polynucleotides, polypeptides and antibodies can be used in	
CC	the detection, diagnosis and treatment (including gene therapy) of	

CC a broad range of pathological states. Clone AC012614.1.0.123 has
CC similarity to cell adhesion molecules, follistatin, roundabout and
CC frazzled, which are involved in neuronal development and
CC reproductive physiology. It may therefore be useful in the
CC treatment of nerve trauma, neurodegenerative disorders, epilepsy,
CC mental health conditions, tissue regeneration in vivo and in
CC vitro, and female reproductive system disorders and pregnancy.

XX
SQ Sequence 815 AA:

Query Match

Best Local Similarity 15.1%; Score 104.5; DB 21; Length 815;
Matches 25; Conservative 10; Mismatches 27; Indels 7; Gaps 4;

QY 53 ASCEGFVCPQOGSPCDI--WDGPFVCRRCRCKEKEPSF--TCASDGLTYNRCYMDAEC 108
Db 36 ascgkklc-srgsrcvisrktgpcqcleac--rpsypvcgsdgrfyenhcklhraac 92
QY 109 LRGLHLHIV 117
Db 93 llgkrltvl 101

Search completed: February 26, 2002, 01:26:42
Job time: 316 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:36 ; Search time 45.78 Seconds
(without alignments)
52.058 Million cell updates/sec

Title: US-09-819-136-2_COPY_93_157
Perfect score: 390
Sequence: 1 CEGFVCPQGSDDCDINDGP.....RCYMDACLRGLHITVPC 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	28.2	306	1 FSL1_MOUSE	Q62356 mus musculus
2	110	28.2	306	1 FSL1_RAT	Q62632 rattus norv
3	110	28.2	308	1 FSL1_HUMAN	Q12841 homo sapien
4	110	28.2	308	1 FSL1_MACRA	Q994Y0 macaca fasc
5	101	25.9	1959	1 AGR1_RAT	P23504 rattus norv
6	100.5	25.6	1328	1 AGR1_CHICK	P31696 gallus gall
7	100	25.6	1328	1 AGR1_DISOM	Q90404 discopysc o
8	93	23.8	341	1 FSA_XENLA	P31515 xenopus lae
9	89	22.8	344	1 FSA_HORSE	O62650 equus cabal
10	88	22.6	343	1 FSA_CHICK	Q90844 gallus gall
11	87	22.3	338	1 FSA_HUMAN	P19885 homo sapien
12	86	22.1	337	1 FSA_SHEEP	P31514 ovis aries
13	86	22.1	344	1 FSA_MOUSE	P47931 mus musculu
14	86	22.1	344	1 FSA_PIG	P10669 sus scrofa
15	86	22.1	344	1 FSA_RAT	P21674 rattus norv
16	83	21.3	344	1 FSA_BOVIN	P50291 bos taurus
17	79.5	20.4	298	1 SPRC_CHICK	P36577 gallus gall
18	79.5	20.4	298	1 SPRC_COTJA	O93390 coturnix co
19	78	20.0	322	1 FSA_BRARE	O93390 coturnix co
20	75	19.2	186	1 IOVO_COTJA	O93390 coturnix co
21	73	18.7	676	1 ORI_COTJA	P23499 coturnix co
22	73	18.7	1429	1 LIT2_CAEEL	P14585 caenorhabdi
23	72	18.5	264	1 SPRC_CAEEL	P34714 caenorhabdi
24	71.5	18.3	273	1 SPRC_RABIT	P36233 oryctolagus
25	71.5	18.3	302	1 SPRC_MOUSE	P07214 mus musculu
26	71.5	18.3	302	1 SPRC_RAT	P16975 rattus norv
27	71.5	18.3	303	1 SPRC_HUMAN	P09486 homo sapien
28	71.5	18.3	304	1 SPRC_BOVIN	P13213 bos taurus
29	71.5	18.3	1217	1 EGF_MOUSE	P01132 mus musculu
30	71	18.2	86	1 ISK4_MOUSE	O35679 mus musculu
31	71	18.2	619	1 MTN4_HUMAN	O95460 homo sapien
32	71	18.2	634	1 SCI_RAT	P24054 rattus norv
33	69.5	17.8	300	1 SPRC_XENLA	P36378 xenopus lae

ALIGNMENTS

34	69.5	17.8	803	1 ITB1_CHICK	P07228 gallus gall
35	68.5	17.6	54	1 IOVO_CARRE	P05616 carpococcyx
36	68.5	17.6	555	1 DP87_DICDI	O04503 dictyostell
37	67.5	17.3	798	1 ITB1_FELCA	P53713 felis silve
38	67.5	17.3	798	1 ITB1_MOUSE	P09055 mus musculu
39	67.5	17.3	1964	1 NTC4_MOUSE	P31695 mus musculu
40	67.5	17.3	1984	1 YL_DROME	P98163 drosophila
41	67	17.2	820	1 AD29_HUMAN	O9ukf5 homo sapien
42	67	17.2	5376	1 ZAN_MOUSE	O88799 mus musculu
43	66.5	17.1	331	1 FHR4_HUMAN	O92496 homo sapien
44	66.5	17.1	787	1 ITB6_MOUSE	O92019 mus musculu
45	66.5	17.1	790	1 AD30_HUMAN	O9ukf2 homo sapien

RESULT 1
ID FSL1_MOUSE STRAND: PRT: 306 AA.
AC Q62356; Q99J19;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN 1 PRECURSOR (TGF-BETA-INDUCIBLE PROTEIN
DE TSC-36).
GN FSTL1 OR FRP OR TSC36.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039028; PubMed=7901004;
RA Shibanuma M., Mashimo J., Mita A., Kuroki T., Nose K.;
RT Cloning from a mouse osteoblastic cell line of a set of
RT transforming-growth-factor-beta 1-regulated genes, one of which
RT seems to encode a follistatin-related polypeptide.;
RT Eur. J. Biochem. 217:13-19(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/Genbank/DDJ databases.
CC - FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
CC - PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC - SIMILARITY: CONTAINS 1 FOLLISTATIN-LIKE DOMAIN.
CC - SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC - SIMILARITY: CONTAINS 1 VMFC DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch.)
CC
CC EMBL: M91380; AAC37633.1; -
CC EMBL: BC006185; AAH06185.1; -
CC HSSP: P01001; 1BUS.
CC MGD: MGI:102793; Fstl.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR003645; FcI.
CC InterPro: IPR001239; Kazal_inh.
CC InterPro: IPR002350; Kazal.
CC Pfam: PF00036; efhand.2.
CC Pfam: PF00050; kazal.1.
CC PRINTS: PR00290; KAZALINHTR.
CC SMART: SM00274; FcI.1.
CC SMART: SM00280; KAZAL.1.
CC PROSITE: PS00282; KAZAL; FALSE_NEG.
CC PROSITE: PS01208; VMFC; FALSE_NEG.

KW Glycoprotein; Signal; Heparin-binding.
 FT SIGNAL 18
 FT CHAIN 19
 FT DOMAIN 19
 FT DOMAIN 28
 FT DOMAIN 51
 FT DOMAIN 52
 FT DOMAIN 52
 FT DOMAIN 231
 FT DISULFID 52
 FT DISULFID 56
 FT DISULFID 75
 FT CARBOHYD 142
 FT CARBOHYD 142
 FT CARBOHYD 173
 FT CARBOHYD 178
 FT CONFLICT 235
 FT SEQUENCE 306 AA; 34538 MW; 4631070AE818F64 CRC64;
 Query Match 28.2%; Score 110; DB 1; Length 306;
 Best Local Similarity 31.7%; Pred. No. 3.4e-05;
 Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;
 QY 1 CEGFVCPQGGSDCDIMW-GQPVCRRCRCKEPEPFTCASDGLTYNRCYMDAECLEGLH 59
 Db 29 CANVFC-GAGRECAVTEKEGPEPTCLICECKPHKRPVCGSNKTYLNHCELHRDACLTSK 87
 QY 60 LHI 62
 Db 88 IQV 90
 RESULT 2
 FSIL_RAT STANDARD; PRT; 306 AA.
 ID FSIL_RAT
 AC 062632;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLISTATIN-RELATED PROTEIN 1 PRECURSOR.
 GN FSTL1 OR FRP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=95045570; PubMed=7957230;
 RA Zwijnen A., Blockx H., van Arnhem W., Willems J., Franssen L.,
 RA Devos K., Raymackers J., van de Voorde A., Slegers H.;
 RT "Characterization of a rat C6 glioma-secreted follistatin-related
 RT protein (FRP). Cloning and sequence of the human homologue.";
 RL Eur. J. Biochem. 225:937-946(1994).
 CC -1- FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
 CC PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: CONTAINS 1 FOLLISTATIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U06864; AAA66063.1; -
 DR HSSP: P02633; IBOC.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR001239; Kazal_inhb.
 DR InterPro: IPR001007; VWFC.
 DR InterPro: IPR002350; kazal.

DR Pfam: PF00036; efhand; 2.
 DR Pfam: PF00050; kazal; 1.
 DR PRINTS: PR00290; KAZALINHBTR.
 DR SMART: SM00274; FOLN; 1.
 DR SMART: SM00280; KAZAL; 1.
 DR SMART: SM00011; VWFC_def; 1.
 DR PROSITE: PS00282; KAZAL; FALSE_NEG.
 DR PROSITE: PS01208; VWFC; FALSE_NEG.
 KW Glycoprotein; Signal; Heparin-binding.
 FT SIGNAL 18
 FT CHAIN 19
 FT DOMAIN 28
 FT DOMAIN 51
 FT DOMAIN 52
 FT DOMAIN 231
 FT DISULFID 52
 FT DISULFID 56
 FT DISULFID 75
 FT CARBOHYD 142
 FT CARBOHYD 142
 FT CARBOHYD 173
 FT CARBOHYD 178
 FT SEQUENCE 306 AA; 34622 MW; 46E197AC7E0CF3D4 CRC64;
 Query Match 28.2%; Score 110; DB 1; Length 306;
 Best Local Similarity 31.7%; Pred. No. 3.4e-05;
 Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;
 QY 1 CEGFVCPQGGSDCDIMW-GQPVCRRCRCKEPEPFTCASDGLTYNRCYMDAECLEGLH 59
 Db 29 CANVFC-GAGRECAVTEKEGPEPTCLICECKPHKRPVCGSNKTYLNHCELHRDACLTSK 87
 QY 60 LHI 62
 Db 88 IQV 90
 RESULT 3
 FSIL_HUMAN STANDARD; PRT; 308 AA.
 ID FSIL_HUMAN
 AC 012841;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLISTATIN-RELATED PROTEIN 1 PRECURSOR.
 GN FSTL1 OR FRP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95045570; PubMed=7957230;
 RA Zwijnen A., Blockx H., van Arnhem W., Willems J., Franssen L.,
 RA Devos K., Raymackers J., van de Voorde A., Slegers H.;
 RT "Characterization of a rat C6 glioma-secreted follistatin-related
 RT protein (FRP). Cloning and sequence of the human homologue.";
 RL Eur. J. Biochem. 225:937-946(1994).
 CC -1- FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
 CC PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN (BY SIMILARITY).
 CC -----
 DR [3]
 DR SEQUENCE FROM N.A.
 DR TISSUE=Kidney;
 DR Strausberg R.;
 DR Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
 CC PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN (BY SIMILARITY).

```

CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 FOLLISTATIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U06863; AAA66062.1; -
DR EMBL: D89937; BA28702.1; -
DR EMBL: BC000055; AAH00055.1; -
DR HSP: P02633; IBOC.
DR MIM: 605547; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR001239; Kazal_inh1b.
DR InterPro: IPR001007; VMFC.
DR InterPro: IPR002350; Kazal.
DR Pfam: PF00036; efnand; 2.
DR Pfam: PF00050; kazal; 1.
DR PRINTS: PR00290; KAZALINHTR.
DR SMART: SM00274; FOLN; 1.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00011; VMC_def; 1.
DR PROSITE: PS00282; KAZAL; FALSE_NEG.
DR PROSITE: PS01208; VMFC; FALSE_NEG.
KW Glycoprotein; Signal; Heparin-binding.
FT SIGNAL 1 20
FT CHAIN 21 308
FT DOMAIN 30 53 FOLLISTATIN-RELATED PROTEIN 1.
FT DOMAIN 54 98 FOLLISTATIN.
FT DOMAIN 233 287 KAZAL-LIKE.
FT DISULFID 54 84 VMFC.
FT DISULFID 58 77 BY SIMILARITY.
FT DISULFID 66 98 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 308 AA; 34985 MW; BDAC651FAFF24800 CRC64;

Query Match 28.2%; Score 110; DB 1; Length 308;
Best Local Similarity 31.7%; Pred. No. 3.4e-05;
Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;

QY 1 CEEFVCPQGGSDCDIMD-GOPVRCRDRCRKESFTCASDGLTYNRCYMDAENCLRGHL 59
DB 31 CANVFC-GAGRECAVTEKEGPTCLCEQCKPHKRPVCGSNKTYLNHCELHRACLGTGSK 89
QY 60 LHI 62
DB 90 IQV 92

RESULT 4
FSL1_MACFA STANDARD: PRT; 308 AA.
AC 09GYO;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN 1 PRECURSOR.
GN FSL1 OR OCC1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;

```

```

RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Tochtant S.;
RT "Neuronal heterogeneity in macaque neocortex revealed by occ1 mRNA
RT activity-dependent expression."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
CC PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 FOLLISTATIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AB039661; BAB20770.1; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; Kazal.
DR InterPro: IPR001007; VMFC.
DR Pfam: PF00036; efnand; 2.
DR Pfam: PF00050; kazal; 1.
DR SMART: SM00274; FOLN; 1.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00214; VMC; 1.
DR SMART: SM00011; VMC_def; 1.
KW Glycoprotein; Signal; Heparin-binding.
FT SIGNAL 1 20
FT CHAIN 21 308
FT DOMAIN 30 53 FOLLISTATIN-RELATED PROTEIN 1.
FT DOMAIN 54 98 FOLLISTATIN.
FT DOMAIN 233 287 KAZAL-LIKE.
FT DISULFID 54 84 VMFC.
FT DISULFID 58 77 BY SIMILARITY.
FT DISULFID 66 98 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 308 AA; 34999 MW; 4B2836D9CDF535D9 CRC64;

Query Match 28.2%; Score 110; DB 1; Length 308;
Best Local Similarity 31.7%; Pred. No. 3.4e-05;
Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;

QY 1 CEEFVCPQGGSDCDIMD-GOPVRCRDRCRKESFTCASDGLTYNRCYMDAENCLRGHL 59
DB 31 CANVFC-GAGRECAVTEKEGPTCLCEQCKPHKRPVCGSNKTYLNHCELHRACLGTGSK 89
QY 60 LHI 62
DB 90 IQV 92

RESULT 5
AGRL_RAT STANDARD: PRT; 1959 AA.
AC P25304; 063034;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AGRN PRECURSOR.
GN AGRN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

NCBI_TaxID=10116;
[1]
RN SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RP TISSUE-Embryonic spinal cord;
RC MEDLINE=9122570; PubMed=1851019;
RA Rupp F., Payan D.G., Magill-Sole C., Cowan D.M., Scheller R.H.;
RT "Structure and expression of a rat agrin.";
RN Neuron 6:811-823(1991).
[2]
RP SEQUENCE OF 1777-1801 FROM N.A.
RX MEDLINE=92407628; PubMed=132608;
RA Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
RT "Structure and chromosomal localization of the mammalian agrin gene.";
RL J Neurosci. 12:3535-3544(1992).
CC -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
CC -1- SUBUNIT: BINDS TO LAMININ.
CC -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
JUNCTION.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
ACETYLCHOLINE RECEPTOR CLUSTERING ACTIVITY.
CC -1- TISSUE SPECIFICITY: EMBRYONIC NERVOUS SYSTEM AND MUSCLE.
CC -1- DEVELOPMENTAL STAGE: MORE ABUNDANT EARLY IN DEVELOPMENT.
CC -1- PTM: CONTAINS HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-
LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 9 KAZAL-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-18 OR MET-24 IS THE
INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: M64780; AAA40703.1; -
DR EMBL: M64780; AAA40702.1; ALT_INIT.
DR EMBL: S44194; AAB23326.1; -
DR PIR: JH0399; AGRT.
DR HSP: P00740; IYXA.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003884; FacI_MAC.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR001239; Kazal_inhib.
DR InterPro: IPR002049; laminin_EGF.
DR InterPro: IPR001791; laminin_G.
DR InterPro: IPR000882; SEA.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00050; kazal; 9.
DR Pfam: PF00053; laminin_EGF; 2.
DR Pfam: PF00054; laminin_G; 3.
DR Pfam: PF01390; SEA; 1.
DR PRINTS: PRO0290; KAZALINHTR.
DR SMART: SM00180; EGF_Lam; 2.
DR SMART: SM00001; EGF_like; 4.
DR SMART: SM00057; FIMAC; 3.
DR SMART: SM00274; FOLN; 5.
DR SMART: SM00280; KAZAL; 9.
DR SMART: SM00282; lamg; 3.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00022; EGF_1; 6.
DR PROSITE: PS01166; EGF_2; 1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE: PS50024; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
Laminin EGF-like domain; Proteoglycan; Heparan sulfate.

FT	SIGNAL	1	29	POTENTIAL.
FT	CHAIN	30	1959	AGRIN.
FT	DOMAIN	65	137	KAZAL-LIKE 1.
FT	DOMAIN	141	212	KAZAL-LIKE 2.
FT	DOMAIN	213	284	KAZAL-LIKE 3.
FT	DOMAIN	287	356	KAZAL-LIKE 4.
FT	DOMAIN	361	429	KAZAL-LIKE 5.
FT	DOMAIN	430	494	KAZAL-LIKE 6.
FT	DOMAIN	495	559	KAZAL-LIKE 7.
FT	DOMAIN	563	645	KAZAL-LIKE 8.
FT	DOMAIN	688	741	LAMININ EGF-LIKE 1.
FT	DOMAIN	742	788	LAMININ EGF-LIKE 2.
FT	DOMAIN	794	864	KAZAL-LIKE 9.
FT	DOMAIN	1023	1145	SEA.
FT	DOMAIN	1220	1258	EGF-LIKE 1.
FT	DOMAIN	1440	1477	EGF-LIKE 2.
FT	DOMAIN	1479	1516	EGF-LIKE 3.
FT	DOMAIN	1709	1748	EGF-LIKE 4.
FT	DOMAIN	869	992	SER/THR-RICH.
FT	DOMAIN	1147	1215	SER/THR-RICH.
FT	DISULFID	97	116	POTENTIAL.
FT	DISULFID	105	137	POTENTIAL.
FT	DISULFID	171	191	POTENTIAL.
FT	DISULFID	180	212	POTENTIAL.
FT	DISULFID	244	263	POTENTIAL.
FT	DISULFID	252	284	POTENTIAL.
FT	DISULFID	316	335	POTENTIAL.
FT	DISULFID	324	356	POTENTIAL.
FT	DISULFID	389	408	POTENTIAL.
FT	DISULFID	397	429	POTENTIAL.
FT	DISULFID	454	473	POTENTIAL.
FT	DISULFID	462	494	POTENTIAL.
FT	DISULFID	518	538	POTENTIAL.
FT	DISULFID	527	559	POTENTIAL.
FT	DISULFID	604	624	POTENTIAL.
FT	DISULFID	613	645	POTENTIAL.
FT	DISULFID	688	700	BY SIMILARITY.
FT	DISULFID	690	707	BY SIMILARITY.
FT	DISULFID	709	718	BY SIMILARITY.
FT	DISULFID	721	739	BY SIMILARITY.
FT	DISULFID	742	754	BY SIMILARITY.
FT	DISULFID	744	761	BY SIMILARITY.
FT	DISULFID	763	772	BY SIMILARITY.
FT	DISULFID	775	786	BY SIMILARITY.
FT	DISULFID	823	843	POTENTIAL.
FT	DISULFID	832	864	POTENTIAL.
FT	DISULFID	1224	1235	BY SIMILARITY.
FT	DISULFID	1229	1246	BY SIMILARITY.
FT	DISULFID	1248	1257	BY SIMILARITY.
FT	DISULFID	1444	1455	POTENTIAL.
FT	DISULFID	1449	1465	POTENTIAL.
FT	DISULFID	1467	1476	POTENTIAL.
FT	DISULFID	1483	1494	BY SIMILARITY.
FT	DISULFID	1488	1504	BY SIMILARITY.
FT	DISULFID	1506	1515	BY SIMILARITY.
FT	DISULFID	1713	1727	BY SIMILARITY.
FT	DISULFID	1721	1736	BY SIMILARITY.
FT	DISULFID	1738	1747	BY SIMILARITY.
FT	CARBOHYD	672	672	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	827	827	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	957	957	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	VARSPPLIC	1144	1152	MISSING (IN ISOFORM 2).
FT	VARSPPLIC	1780	1798	MISSING (IN ISOFORM 3).
FT	VARSPPLIC	1788	1798	MISSING (IN ISOFORM 4).
FT	VARSPPLIC	1780	1787	MISSING (IN ISOFORM 5).
FT	VARSPPLIC	314	314	V -> VTCD (IN A VARIANT).
FT	SEQUENCE	1959	208645	MM; 7FERDFDAFF89CC31 CRC64;

Query Match 25.9%; Score 101; DB 1; Length 1959;
Best Local Similarity 30.9%; Pred. NO. 0.0016;
Matches 21; Conservative 12; Mismatches 31; Indels 4; Gaps 2;

OY 1 CEGFVCPGSGDIDWDGPVCRCHDRCKEPEFTFCASDGLTYNNRCYMDAEACLRGLHL 60
DB 363 CAGVGC-ARFAGVCTVKNKAECCEQKVCSTGYDPVCGSDGVYGVCELSMACLTGREI 421
OY 61 HIV---PC 65
DB 422 QVARRGPC 429

RESULT 6
AGRI_CHICK STANDARD; PRT; 1955 AA.
ID AGRI_CHICK
AC P31696;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AGRIN PRECURSOR.
GN AGRN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92232297; PubMed=1314620;
RA Tsim K.W.K., Ruegg M.A., Escher G., Kroege S., McMahon U.J.;
RT "CDNA that encodes active agrin.";
RL Neuron 8:677-689(1992).
RN [2]
RP ALTERNATIVE SPLICING.
RX MEDLINE=92232298; PubMed=1314621;
RA Ruegg M.A., Tsim K.W.K., Horton S.E., Kroege S., Escher G.,
RA Gensch E.M., McMahon U.J.;
RT "The agrin gene codes for a family of basal lamina proteins that
RT differ in function and distribution.";
RL Neuron 8:691-699(1992).
CC -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
CC -1- SUBUNIT: BINDS TO LAMININ.
CC -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
CC JUNCTION.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE DIFFERENT FORMS ARISE BY
CC ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR
CC CLUSTERING ACTIVITY.
CC -1- PWM: CONTAINS HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-
CC LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 9 KAZAL-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: M94571; AAA48585.1; -;
DR EMBL: M97371; AAA48586.1; -;
DR EMBL: M97372; -; NOT_ANNOTATED_CDS.
DR PIR: JH0591; AGCH.
DR InterPro: IPR000152; ASX_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR003884; FACI_MAC.
DR InterPro: IPR003645; FOIN.
DR InterPro: IPR001239; Kazal_inhib.

DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR002350; Kazal.
DR Pfam: PF00008; EGF_4.
DR Pfam: PF00050; Kazal; 9.
DR Pfam: PF00053; Laminin_EGF; 2.
DR Pfam: PF00054; Laminin_G; 3.
DR Pfam: PF01390; SEA; 1.
DR PRINTS: PRO0290; KAZALINHTR.
DR SMART: SM00180; EGF_Lam; 2.
DR SMART: SM00001; EGF_Like; 4.
DR SMART: SM00057; FIMAC; 2.
DR SMART: SM00274; POLN; 5.
DR SMART: SM00280; KAZAL; 9.
DR SMART: SM00282; LamG; 3.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; 6.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01248; LAMININ_Type_EGF; 1.
DR PROSITE: PS50024; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT SIGNAL 1 38
FT CHAIN 1 1955
FT FT 39 1955 AGRIN.
FT FT 54 126 KAZAL-LIKE 1.
FT FT 130 201 KAZAL-LIKE 2.
FT FT 202 273 KAZAL-LIKE 3.
FT FT 276 344 KAZAL-LIKE 4.
FT FT 350 418 KAZAL-LIKE 5.
FT FT 419 483 KAZAL-LIKE 6.
FT FT 484 548 KAZAL-LIKE 7.
FT FT 551 633 KAZAL-LIKE 8.
FT FT 675 728 LAMININ EGF-LIKE 1.
FT FT 729 775 LAMININ EGF-LIKE 2.
FT FT 781 851 KAZAL-LIKE 9.
FT FT 856 995 SER/THR-RICH.
FT FT 1150 1219 SER/THR-RICH.
FT FT 1229 1265 EGF-LIKE 1.
FT FT 1446 1483 EGF-LIKE 2.
FT FT 1485 1522 EGF-LIKE 3.
FT FT 1714 1752 EGF-LIKE 4.
FT FT DOMAIN 86 105 POTENTIAL.
FT FT DISULFID 94 126 POTENTIAL.
FT FT DISULFID 160 180 POTENTIAL.
FT FT DISULFID 169 201 POTENTIAL.
FT FT DISULFID 233 252 POTENTIAL.
FT FT DISULFID 241 273 POTENTIAL.
FT FT DISULFID 304 323 POTENTIAL.
FT FT DISULFID 312 344 POTENTIAL.
FT FT DISULFID 378 397 POTENTIAL.
FT FT DISULFID 386 418 POTENTIAL.
FT FT DISULFID 443 462 POTENTIAL.
FT FT DISULFID 451 483 POTENTIAL.
FT FT DISULFID 507 527 POTENTIAL.
FT FT DISULFID 516 548 POTENTIAL.
FT FT DISULFID 552 612 POTENTIAL.
FT FT DISULFID 601 633 POTENTIAL.
FT FT DISULFID 675 687 BY SIMILARITY.
FT FT DISULFID 677 694 BY SIMILARITY.
FT FT DISULFID 696 705 BY SIMILARITY.
FT FT DISULFID 708 726 BY SIMILARITY.
FT FT DISULFID 729 741 BY SIMILARITY.
FT FT DISULFID 731 748 BY SIMILARITY.
FT FT DISULFID 750 759 BY SIMILARITY.
FT FT DISULFID 762 773 BY SIMILARITY.
FT FT DISULFID 810 830 POTENTIAL.
FT FT DISULFID 819 851 POTENTIAL.
FT FT DISULFID 1233 1244 BY SIMILARITY.
FT FT DISULFID 1238 1253 BY SIMILARITY.
FT FT DISULFID 1255 1264 BY SIMILARITY.
FT FT DISULFID 1450 1461 BY SIMILARITY.

FT DISULFID 1455 1471 BY SIMILARITY.
 FT DISULFID 1473 1482 BY SIMILARITY.
 FT DISULFID 1489 1500 BY SIMILARITY.
 FT DISULFID 1494 1510 BY SIMILARITY.
 FT DISULFID 1512 1521 BY SIMILARITY.
 FT DISULFID 1718 1731 BY SIMILARITY.
 FT DISULFID 1725 1740 BY SIMILARITY.
 FT DISULFID 1742 1751 BY SIMILARITY.
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 659 659 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1648 1651 MISSING (IN AGRIN-RELATED PROTEIN 2).
 FT VARSPLIC 1783 1793 MISSING (IN AGRIN-RELATED PROTEIN 1 AND AGRIN-RELATED PROTEIN 2).
 FT CONFLICT 1129 1131 RTI -> STL (IN REF. 1; AAA48586).
 FT SEQUENCE 1955 AA; 211411 MW; B4DB27C23422581 CRC64;

Query Match 25.8%; Score 100.5; DB 1; Length 1955;
 Best Local Similarity 35.6%; Pred. No. 0.0018;
 Matches 21; Conservative 11; Mismatches 24; Indels 3; Gaps 2;

Oy 10 GSDCINMGQPCRCRDRCEKESFTCASDGLTYNRCYMAEACLT--RGHLH-HIVPC 65
 Db 360 GATCVAKNRPVCECOVCGRHYDPCGSDNRTYGNPCELNMACVLKREIKVKGKGPC 418

RESULT 7
 AGRI_DISOM STANDARD; PRT: 1328 AA.
 ID AGRI_DISOM
 AC 090404;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AGRIN (FRAGMENT).
 GN AGRN.
 OS discopoge ommata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalae; Hypnosqualae; Pristiogaster; Batoidae;
 OC Torpediniformes; Narchnoidei; Narchnidae; Discopoge.
 ON NCBI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith M.A., Magill J., Solic C., Rupp F., Yao Y.-M.M., Schilling J.W.,
 RA Show P., McMahon U.J.;
 RT "Isolation and characterization of a cDNA that encodes an agrin
 homolog in the marine ray."
 RL Mol. Cell. Neurosci. 3:406-417(1992).
 CC -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
 CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
 CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
 CC JUNCTION (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS AT LEAST 2 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS AT LEAST 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L01423; AAA49224.1; -
 DR HSSP; P00740; IIXA.
 DR InterPro: IPR000153; Asx hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR003884; FacI_MAC.

DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR000082; SEA.
 DR InterPro: IPR002350; kazal.
 DR Pfam: PF00008; EGF; 4.
 DR Pfam: PF00050; kazal; 2.
 DR Pfam: PF00053; Laminin_EGF; 2.
 DR Pfam: PF00054; Laminin_G; 3.
 DR Pfam: PF01390; SEA; 1.
 DR SMART: SM00180; EGF_Lam; 2.
 DR SMART: SM00001; EGF_Like; 3.
 DR SMART: SM00057; FYAC; 1.
 DR SMART: SM00274; FOLN; 1.
 DR SMART: SM00280; KAZAL; 2.
 DR SMART: SM00282; LamG; 3.
 DR SMART: SM00200; SEA; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; 5.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE: PS50024; SEA; 1.
 DR Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
 FT NON_TER 1 1
 FT DOMAIN 79 132 LAMININ EGF-LIKE 1.
 FT DOMAIN 133 179 LAMININ EGF-LIKE 2.
 FT DOMAIN 411 533 SEA.
 FT DOMAIN 608 644 EGF-LIKE 1.
 FT DOMAIN 865 902 EGF-LIKE 2.
 FT DOMAIN 1097 1135 EGF-LIKE 3.
 FT DISULFID 81 98 BY SIMILARITY.
 FT DISULFID 100 109 BY SIMILARITY.
 FT DISULFID 112 130 BY SIMILARITY.
 FT DISULFID 133 145 BY SIMILARITY.
 FT DISULFID 135 152 BY SIMILARITY.
 FT DISULFID 154 163 BY SIMILARITY.
 FT DISULFID 166 177 BY SIMILARITY.
 FT DISULFID 612 623 BY SIMILARITY.
 FT DISULFID 617 632 BY SIMILARITY.
 FT DISULFID 634 643 BY SIMILARITY.
 FT DISULFID 869 880 BY SIMILARITY.
 FT DISULFID 874 890 BY SIMILARITY.
 FT DISULFID 892 901 BY SIMILARITY.
 FT DISULFID 1101 1114 BY SIMILARITY.
 FT DISULFID 1108 1123 BY SIMILARITY.
 FT DISULFID 1125 1134 BY SIMILARITY.
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1328 AA; 144018 MW; 79D81C1AF2A71C18 CRC64;

Query Match 25.6%; Score 100; DB 1; Length 1328;
 Best Local Similarity 35.7%; Pred. No. 0.0014;
 Matches 25; Conservative 8; Mismatches 31; Indels 6; Gaps 4;

Oy 1 CEEFVPCQSSDDINDGVCRCRDR-CEKESF--TCASDGLTYNRCYMAEACLTREGI 58
 Db 187 GSDLC-H-QGATVCQVSGISGRVCEPCPSICPKNKGVCSDGVYANECOLKTIACRQGS 245
 Oy 59 HLHIV--PC 65
 Db 246 VINILHQGPC 255
 RESULT 8
 FSA_XENLA STANDARD; PRT: 341 AA.
 ID FSA_XENLA
 AC P31515; 091376;
 DT 01-JUL-1993 (Rel. 26, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN) (XFS-319).
 GN FST.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 CC Xenopodinae; Xenopus.
 NC NCBITaxID=8335;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND CHARACTERIZATION.
 RC TISSUE-Tail bud;
 RX MEDLINE=94221645; PubMed=8168135;
 RA Hemmati-Brivanlou A., Kelly O.G., Melton D.A.;
 RT "Follistatin, an antagonist of activin, is expressed in the Spemann
 RT organizer and displays direct neutralizing activity.";
 RL Cell 77:283-295(1994).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A. (LONG ISOFORM).
 RC TISSUE-Ovary;
 RX MEDLINE=91128357; PubMed=1704219;
 RA Tashiro K., Yamada R., Asano M., Hashimoto M., Muramatsu M.,
 RA Shiohara K.;
 RT "Expression of mRNA for activin-binding protein (follistatin) during
 RT early embryonic development of Xenopus laevis.";
 RL Biochem. Biophys. Res. Commun. 174:1022-1027(1991).
 RN [3]
 RP SEQUENCE OF N-TERMINUS, AND CHARACTERIZATION.
 RX PubMed=8365557;
 RA Fukui A., Nakamura T., Sugino K., Uchiyama H., Asashima M.,
 RA Sugino H.;
 RT "Isolation and characterization of Xenopus follistatin and activins.";
 RL Dev. Biol. 159:131-139(1993).
 CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 CC ANTAGONIST WHICH PLAYS A ROLE IN NEURAL INDUCTION. THE SHORT
 CC ISOFORM IS A MORE POTENT INHIBITOR OF ACTIVIN THAN THE LONG
 CC ISOFORM. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION OF
 CC PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: SPERMANN ORGANIZER AND NOTOCHORD.
 CC -1- DEVELOPMENTAL STAGE: THE SHORT ISOFORM IS PRESENT MATERNAALLY WHILE
 CC THE LONG ISOFORM IS EXPRESSED AT GASTRULA STAGES. DETECTED IN A
 CC FEW CELLS OF THE SPERMANN ORGANIZER AT THE ONSET OF GASTRULATION.
 CC DURING GASTRULATION EXPRESSION CONTINUES IN THE PRECHORDAL PLATE
 CC AND THE ANTERIOR PORTION OF THE NOTOCHORD ANLAGE. BEGINNING AT
 CC EARLY NEURULA STAGES, EXPRESSION IS INITIATED AT NEW SITES IN THE
 CC HEAD MESODERM; HYPOCHORD; PRONEPHROS. EYES, FORE-, MID-, AND
 CC HINDBRAIN; AND THE MIDBRAIN-HINDBRAIN JUNCTION.
 CC -1- INDUCTION: BY ACTIVIN.
 CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 DR EMBL; S69801; AAB30638.1; .
 DR PIR; JN0119; JN0119.
 DR HSSP; P37109; IPCE.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR002350; KAZAL.
 DR Pfam; PF00050; KAZAL; 3.
 DR SMART; SM00274; FOLN; 2.
 DR SMART; SM00280; KAZAL; 3.
 DR PROSITE; PS00282; KAZAL; FALSE NEG.
 KW Glycoprotein; Repeat; Signal; Alternative splicing.

FT SIGNAL 1 29
 FT CHAIN 30 341 FOLLISTATIN.
 FT DOMAIN 94 117 FOLLISTATIN 1.
 FT DOMAIN 118 164 KAZAL-LIKE 1.
 FT DOMAIN 167 190 FOLLISTATIN 2.
 FT DOMAIN 192 239 KAZAL-LIKE 2.
 FT DOMAIN 244 268 FOLLISTATIN 3.
 FT DOMAIN 270 316 FOLLISTATIN 3.
 FT DOMAIN 324 331 KAZAL-LIKE 3.
 FT DOMAIN 118 150 POLY-GLU.
 FT DISULFID 122 143 BY SIMILARITY.
 FT DISULFID 132 164 BY SIMILARITY.
 FT DISULFID 192 225 BY SIMILARITY.
 FT DISULFID 196 218 BY SIMILARITY.
 FT DISULFID 207 239 BY SIMILARITY.
 FT DISULFID 270 302 BY SIMILARITY.
 FT DISULFID 274 295 BY SIMILARITY.
 FT DISULFID 284 316 BY SIMILARITY.
 FT CARBOHYD 72 72 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT VARSPLIC 318 341 SIVEDTEEEEEEEDYFVTSW -> CK (IN SHORT
 FT CONFLICT 268 268 A -> G (IN REF. 2).
 FT CONFLICT 270 270 C -> S (IN REF. 2).
 SQ SEQUENCE 341 AA; 37543 MW; 5D26E7BE70890171 CRC64;

 Query Match 23.8%; Score 93; DB 1; Length 341;
 Best Local Similarity 36.8%; Pred. No. 0.0027;
 Matches 25; Conservative 5; Mismatches 30; Indels 8; Gaps 5;

 QY 1 CEGFVCPQGGSDCDIMD---GQPVCR-CDRRC--EKEPSFTCASDGLTYNRCYMAENC 54
 DB 245 CEDIQ-SAGKRC-LWDSRVGRKCALCDLGGESKSDPTVCASDNTYPSBCAMQAC 302
 QY 55 LRLGLHLHI 62
 DB 303 STGILLEV 310

 RESULT 9
 FSA_HORSE STANDARD; PRT; 344 AA.
 AC 062550;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
 GN FST.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 NC NCBITaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Follicle;
 RA Sugawara Y., Yamanouchi K.;
 RL Submitter (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 CC OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>

```

CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: AB010829; BAA25699.1;
CC InterPro: IPR003645; FOLN.
CC InterPro: IPR002350; kazal.
CC Pfam: PF00050; kazal.3.
CC SMART: SM00274; FOLN; 3.
CC SMART: SM00280; KAZAL; 3.
CC DR PROSITE: PS00282; KAZAL; FALSE_NEG.
CC KW Glycoprotein; Repeat; Signal.
CC FT SIGNAL 1 29
CC FT CHAIN 30 344
CC FT DOMAIN 94 117
CC FT DOMAIN 118 164
CC FT DOMAIN 167 190
CC FT DOMAIN 192 239
CC FT DOMAIN 244 268
CC FT DOMAIN 270 316
CC FT DISULFID 118 150
CC FT DISULFID 122 143
CC FT DISULFID 132 164
CC FT DISULFID 192 225
CC FT DISULFID 196 218
CC FT DISULFID 207 239
CC FT DISULFID 270 302
CC FT DISULFID 274 295
CC FT DISULFID 284 316
CC FT CARBOHYD 124 124
CC FT CARBOHYD 288 288
CC SQ SEQUENCE 344 AA; 38002 MW; 1803577D2D9BE4AA CRC64;

Query Match 22.8%; Score 89; DB 1; Length 344;
Best Local Similarity 35.7%; Pred. NO. 0.0073;
Matches 25; Conservative 6; Mismatches 27; Indels 12; Gaps 5;

QY 1 CEGVCPQGGSDCDIMD---GQPVCR-----CRDRCRCEPFTGASDGLTYNNCYMDAE 52
Db 245 CEDTQC-TGKKKC-LMPFKYGRGRCSLCDELCPDPSKEP--VCASDNATYVSECANKEA 300
QY 53 AGLRGLHHT 62
Db 301 ACSSGVLLLEV 310

RESULT 10
FSA_CHICK STANDARD; PRT; 343 AA.
AC Q90844; Q9P597; Q90680;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
GN FST.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96189336; PubMed=8625798;
RA Graham A., Lumsden A.;
RT "Interactions between rhombomeres modulate Krox-20 and follistatin
RL expression in the chick embryo hindbrain.";
RL Development 122:473-480(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96029008; PubMed=7554497;
RA Connolly D.J., Patel K., Seletro E.A., Wilkinson D.G., Cooke J.;
RT "Cloning, sequencing, and expression analysis of the chick homologue

```

```

RT of follistatin".
RL Dev. Genet. 17:65-77(1995).
RN [3]
RP SEQUENCE OF 177-250 FROM N.A.
RC TISSUE=Embryonic ovary;
RX MEDLINE=96049133; PubMed=7576634;
RA Darland D.C., Link B.A., Nishi R.;
RT "Activin A and follistatin expression in developing targets of ciliary
RT ganglion neurons suggests a role in regulating neurotransmitter
RT phenotype.";
RL Neuron 15:857-866(1995).
CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
CC ANTAGONIST. INHIBITS ACTIVIN A SIGNALING IN THE IRIS AND REGULATES
CC SOMATOSTATIN PHENOTYPE IN CILIARY GANGLION NEURONS. SPECIFIC
CC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION OF PITUITARY FOLLICLE
CC STIMULATING HORMONE (FSH).
CC -1- SUBUNIT: MONOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: CILIARY GANGLION NEURONS. LEVELS ARE HIGHER IN
CC THE IRIS THAN THE CHOROID.
CC -1- DEVELOPMENTAL STAGE: LEVELS INCREASE IN THE IRIS FROM EMBRYONIC
CC DAY 9 (E9) TO E16 IN CONTRAST TO THE CHOROID WHERE IT REMAINS LOW
CC RELATIVE TO IRIS. DURING EARLY HINDBRAIN DEVELOPMENT STRONGLY
CC EXPRESSED IN RHOMBOMERES R2, R4, R5 AND R6 BUT NOT IN R3.
CC EXPRESSION IN R3 IS SEEN AT LATER STAGES AND IS DEPENDENT ON
CC NEIGHBOURING INTERACTIONS.
CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: X87609; CAA60915.1;
CC DR EMBL: U34589; AAA92005.1;
CC DR HSSP: P00998; ITGS.
CC DR InterPro: IPR003645; FOLN.
CC DR InterPro: IPR002350; kazal.
CC DR Pfam: PF00050; kazal; 3.
CC DR SMART: SM00274; FOLN; 3.
CC DR SMART: SM00280; KAZAL; 3.
CC DR PROSITE: PS00282; KAZAL; FALSE_NEG.
CC DR KW Glycoprotein; Repeat; Signal.
CC FT SIGNAL 1 28
CC FT CHAIN 29 343
CC FT DOMAIN 93 343
CC FT DOMAIN 117 163
CC FT DOMAIN 166 189
CC FT DOMAIN 191 238
CC FT DOMAIN 243 267
CC FT DOMAIN 269 315
CC FT DOMAIN 323 329
CC FT DISULFID 117 149
CC FT DISULFID 121 142
CC FT DISULFID 131 163
CC FT DISULFID 191 224
CC FT DISULFID 195 217
CC FT DISULFID 206 238
CC FT DISULFID 269 301
CC FT DISULFID 273 294
CC FT DISULFID 283 315
CC FT CARBOHYD 123 123
CC FT CARBOHYD 267 287
CC FT CARBOHYD 129 129
CC FT CARBOHYD 177 177
CC FT CONFLICT 183 183
CC FT CONFLICT 202 202
CC FT CONFLICT 228 228
CC SQ SEQUENCE 343 AA; 38192 MW; 15A78762360E1AAE CRC64;

```

DOMAIN	196	239	NAME-PLATE 2.
F1			

DR SMART; SM00274; FOLN; 3.
 DR SMART; SM00280; KAZAL; 3.
 DR PROSITE; PS00282; KAZAL; FALSE_NEG.
 KW Glycoprotein; Repeat; Signal.
 FT NON_TER 1
 FT SIGNAL <1 22
 FT CHAIN 23 337 FOLLISTATIN.
 FT DOMAIN 87 110 FOLLISTATIN 1.
 FT DOMAIN 111 157 KAZAL-LIKE 1.
 FT DOMAIN 160 183 FOLLISTATIN 2.
 FT DOMAIN 185 232 KAZAL-LIKE 2.
 FT DOMAIN 237 261 FOLLISTATIN 3.
 FT DOMAIN 263 309 KAZAL-LIKE 3.
 FT DOMAIN 314 326 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DISULFID 111 143 BY SIMILARITY.
 FT DISULFID 115 136 BY SIMILARITY.
 FT DISULFID 125 157 BY SIMILARITY.
 FT DISULFID 185 218 BY SIMILARITY.
 FT DISULFID 189 211 BY SIMILARITY.
 FT DISULFID 200 232 BY SIMILARITY.
 FT DISULFID 263 295 BY SIMILARITY.
 FT DISULFID 267 288 BY SIMILARITY.
 FT DISULFID 277 309 BY SIMILARITY.
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 337 AA; 37082 MW; 1E8B1BBB6B109C4 CRC64;

Query Match 22.1%; Score 86; DB 1; Length 337;
 Best Local Similarity 35.7%; Pred. No. 0.015;
 Matches 25; Conservative 7; Mismatches 26; Indels 12; Gaps 6;

QY 1 CEEFVCPQGSDDCDIWD---GQPYCR-CRDRC---EKESPTGASDGLTYNRCYMDAE 52
 DB 238 CEDIQ-C-TGSKKC-LMDPKVGRGRCSLCDLCPDSKSEEP--VCASDNATYASCCAMKEA 293
 OY 53 ACLRGHLHI 62
 DB 294 ACSSGVLEEV 303

RESULT 13
 FSA_MOUSE /
 ID FSA_MOUSE STANDARD; PRT; 344 AA.
 AC P47931;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
 GN FST.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE=93324368; PubMed=7600958;
 RA Albano R.M., Arkell R., Beedington R.S.P., Smith J.C.;
 RT *Expression of inhibin subunits and follistatin during
 RT postimplantation mouse development: decidual expression of activin
 RT and expression of follistatin in primitive streak, somites and
 RT hindbrain.*;
 RT hindbrain.*;
 RL Development 120:803-813(1994).
 RN [2]
 RP SEQUENCE OF X-340 FROM N.A.
 RC STRAIN=CBA X NMRI; TISSUE-Ovary;
 RX MEDLINE=95045246; PubMed=7956942;
 RA Tuuri T., Ermaa M., Hilden K., Riltvos O.;
 RT *Activin-binding protein follistatin messenger ribonucleic acid and
 RT secreted protein levels are induced by chorionic gonadotropin in
 RT cultured human granulosa-luteal cells.*;
 RT Endocrinology 135:2196-2203(1994).
 RL

CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 CC OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DEVELOPMENTAL STAGE: EMBRYONIC EXPRESSION FIRST OCCURS IN THE
 CC PRIMITIVE STREAK, FOLLOWED BY EXPRESSION IN HEAD MESODERM,
 CC SOMITES, AND SPECIFIC RHOMBOMERES OF THE HINDBRNIN AND LATER IN
 CC MIDBRNIN AND DIENCEPHALON. NO EXPRESSION IS SEEN IN THE NODE OR
 CC NOTOCHORD.
 CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z29532; CA82648.1; -;
 DR EMBL; X83377; CA58291.1; -;
 DR HSP; P00998; ITGS.
 DR MCD; MGI:95586; FSL.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR002350; kazal.
 DR Pfam; PF00050; kazal; 3.
 DR SMART; SM00274; FOLN; 3.
 DR SMART; SM00280; KAZAL; 3.
 DR PROSITE; PS00282; KAZAL; FALSE_NEG.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 344 POTENTIAL.
 FT DOMAIN 94 117 FOLLISTATIN 1.
 FT DOMAIN 118 164 KAZAL-LIKE 1.
 FT DOMAIN 167 190 FOLLISTATIN 2.
 FT DOMAIN 192 239 KAZAL-LIKE 2.
 FT DOMAIN 244 268 FOLLISTATIN 3.
 FT DOMAIN 270 316 KAZAL-LIKE 3.
 FT DISULFID 118 150 BY SIMILARITY.
 FT DISULFID 122 143 BY SIMILARITY.
 FT DISULFID 132 164 BY SIMILARITY.
 FT DISULFID 192 225 BY SIMILARITY.
 FT DISULFID 196 218 BY SIMILARITY.
 FT DISULFID 207 239 BY SIMILARITY.
 FT DISULFID 270 302 BY SIMILARITY.
 FT DISULFID 274 295 BY SIMILARITY.
 FT DISULFID 284 316 BY SIMILARITY.
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 241 242 KA -> T (IN REF. 1).
 SQ SEQUENCE 344 AA; 37866 MW; 935B6CBB213176F9 CRC64;

Query Match 22.1%; Score 86; DB 1; Length 344;
 Best Local Similarity 35.7%; Pred. No. 0.016;
 Matches 25; Conservative 5; Mismatches 28; Indels 12; Gaps 5;

QY 1 CEEFVCPQGSDDCDIWD---GQPYCR-CRDRCEKPSFTGASDGLTYNRCYMDAE 52
 DB 245 CEDIQ-GGSKKC-LMDPKVGRGRCSLCDLCPDSKSEEP--VCASDNATYASCCAMKEA 300
 OY 53 ACLRGHLHI 62
 DB 301 ACSSGVLEEV 310

RESULT 14
 FSA_PIG /
 ID FSA_PIG STANDARD; PRT; 344 AA.
 AC P10669; P10670;
 DT 01-JUL-1989 (Rel. 11, Created)

01-JUL-1989 (Rel. 11, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
 GN FST.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88209050; PubMed=3365249;
 RA Shimazaki S., Koga M., Esch F., Mercado M., Cooksey K., Koba A.,
 Ling N.;
 RT "Porcine follistatin gene structure supports two forms of mature
 follistatin produced by alternative splicing."
 RL Biochem. Biophys. Res. Commun. 152:717-723(1988).
 RN [2]
 RP SEQUENCE OF 1-334 FROM N.A.
 RX MEDLINE=91042571; PubMed=3153465;
 RA Esch F.S., Shimazaki S., Mercado M., Cooksey K., Ling N., Ying S.,
 Ueno N., Guillemain R.;
 RT "Structural characterization of follistatin: a novel follicle-
 stimulating hormone release-inhibiting polypeptide from the gonad."
 RL Mol. Endocrinol. 1:849-855(1987).
 CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND A'; ARE
 PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M19529; AAA31036.1; -
 DR EMBL: M19529; AAA31037.1; -
 DR EMBL: M36512; AAA31038.1; -
 DR PIR: A27701; A27701.
 DR HSSP: P00998; 1TGS.
 DR InterPro: IPR003645; FOLN.
 DR Pfam: PF0002350; kazal.
 DR SMART: SM00274; FOLN; 3.
 DR SMART: SM00280; KAZAL; 3.
 DR PROSITE: PS00282; KAZAL; FALSE_NEG.
 KW Glycoprotein; Repeat; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 344
 FT DOMAIN 94 117 FOLLISTATIN 1.
 FT DOMAIN 118 164 KAZAL-LIKE 1.
 FT DOMAIN 167 190 FOLLISTATIN 2.
 FT DOMAIN 192 239 KAZAL-LIKE 2.
 FT DOMAIN 244 268 FOLLISTATIN 3.
 FT DOMAIN 270 316 KAZAL-LIKE 3.
 FT DOMAIN 321 333 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 150 150 BY SIMILARITY.
 FT DISULFID 122 143 BY SIMILARITY.
 FT DISULFID 132 164 BY SIMILARITY.
 FT DISULFID 192 225 BY SIMILARITY.
 FT DISULFID 196 218 BY SIMILARITY.
 FT DISULFID 207 239 BY SIMILARITY.
 FT DISULFID 270 302 BY SIMILARITY.
 FT DISULFID 274 295 BY SIMILARITY.
 FT DISULFID 284 316 BY SIMILARITY.
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 318 344 MISSING (IN ISOFORM A').
 SQ SEQUENCE 344 AA; 38035 MW; 6906E7A23C9BD21 CRC64;
 Query Match 22.1%; Score 86; DB 1; Length 344;
 Best local Similarity 35.7%; Pred. No. 0.016;
 Matches 25; Conservative 7; Mismatches 26; Indels 12; Gaps 6;
 QY 1 CEEFVCPQSGSDCDIND---GQPVCR-CRDRC-----EKEPSFYCASDGLTYNRCYVDAE 52
 Db 245 CEDIQ-CGGKRC-LWDFKVGRCRSLCDLCEPSKSEEP--VCASDNATYASECAKREA 300
 QY 53 ACIRGLHLH 62
 Db 301 ACSSGVLEV 310
 RESULT 15
 FSA_RAT STANDARD; PRT; 344 AA.
 AC P21674;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
 GN FST.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE=89261821; PubMed=2725528;
 RA Shimazaki S., Koga M., Buscaglia M.L., Simmons D.M., Bicsak T.A.,
 Ling N.;
 RT "Follistatin gene expression in the ovary and extragonadal tissues."
 RL Mol. Endocrinol. 3:651-659(1989).
 CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M31591; AAB6704.1; -
 DR EMBL: M31586; AAB6704.1; JOINED.
 DR EMBL: M31587; AAB6704.1; JOINED.
 DR EMBL: M31588; AAB6704.1; JOINED.
 DR EMBL: M31589; AAB6704.1; JOINED.
 DR EMBL: M31590; AAB6704.1; JOINED.
 DR HSSP: P00998; 1TGS.
 DR InterPro: IPR003645; FOLN.
 DR Pfam: PF0002350; kazal.
 DR SMART: SM00274; FOLN; 3.
 DR SMART: SM00280; KAZAL; 3.
 DR PROSITE: PS00282; KAZAL; FALSE_NEG.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 344
 FT DOMAIN 94 117 FOLLISTATIN 1.
 FT DOMAIN 118 164 KAZAL-LIKE 1.

```
FT DOMAIN 167 190 FOLLISTATIN 2.  
FT DOMAIN 192 239 KAZAL-LIKE 2.  
FT DOMAIN 244 268 FOLLISTATIN 3.  
FT DOMAIN 270 316 KAZAL-LIKE 3.  
FT DOMAIN 321 333 ASP/GLU-RICH (HIGHLY ACIDIC).  
FT DISULFID 118 150 BY SIMILARITY.  
FT DISULFID 122 143 BY SIMILARITY.  
FT DISULFID 132 164 BY SIMILARITY.  
FT DISULFID 192 225 BY SIMILARITY.  
FT DISULFID 196 218 BY SIMILARITY.  
FT DISULFID 207 239 BY SIMILARITY.  
FT DISULFID 270 302 BY SIMILARITY.  
FT DISULFID 274 295 BY SIMILARITY.  
FT DISULFID 284 316 BY SIMILARITY.  
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 344 AA; 37838 MM; 864244CF05436552 CRC64;
```

```
Query Match 22.1%; Score 86; DB 1; Length 344;  
Best Local Similarity 35.7%; Pred. No. 0.016;  
Matches 25; Conservative 5; Mismatches 28; Indels 12; Gaps 5;
```

```
OY 1 CEGFVCPQGGSDCDIWD---GQPVCR-----CRDRCEKEPSFTCASDGLTYNNRCYMDAE 52  
Db 245 CEDIQG-GGGRKC-LWDRKVRGRGRCISICDELCPDSKSDP--VCASDNATYASQCAMEKA 300  
OY 53 ACLRGLHLHI 62  
Db 301 ACSSGVILLEV 310
```

Search completed: February 26, 2002, 01:46:37
Job time: 1181 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:29:33 ; Search time 78.07 seconds
(without alignments)
63,422 Million cell updates/sec

Title: US-09-819-136-2_COPY_93_157

Perfect score: 390
Sequence: 1 CEGFVCPQSGSDCDIMDGP.....RCYMDAECRLGLHLIVPC 65

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR68:*

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	120	30.8	1296	2 T16859	hypothetical prote
2	110	28.2	299	2 JG0187	follicle-statin-relate
3	110	28.2	306	2 S38251	follicle-statin-relate
4	110	28.2	306	2 S51361	follicle-statin-relate
5	110	28.2	308	2 S51362	follicle-statin-relate
6	101	25.9	1959	1 AGRT	agrin - rat
7	100.5	25.8	1955	1 AGCH	agrin precursor -
8	100	25.6	1328	2 T43060	agrin - electric r
9	93	23.8	319	2 A53502	follicle-statin - Afril
10	91	23.3	773	2 T46283	hypothetical prote
11	88	22.6	343	2 S55369	follicle-statin - chic
12	87	22.3	344	2 A32141	follicle-statin 1 prec
13	86	22.1	337	2 T47079	follicle-statin - shee
14	86	22.1	343	2 S45321	follicle-statin - mous
15	86	22.1	344	1 A27701	follicle-statin precu
16	86	22.1	344	1 I57698	follicle-statin - rat
17	84	21.5	380	2 G01639	transmembrane prot
18	83	21.3	344	2 T45894	follicle-statin - bovl
19	80.5	20.6	298	2 T46914	hypothetical prote
20	79.5	20.4	298	2 S39539	osteonectin precu
21	75	19.2	186	1 T1QJM	ovomucoid - Japane
22	75	19.2	228	1 S45677	proteolase inhibit
23	74	19.0	1162	2 T21557	hypothetical prote
24	73	18.7	676	1 A39379	hatching-supprese
25	73	18.7	1429	2 S05434	homeotic protein 1
26	72	18.5	264	1 A47737	osteonectin precu
27	71.5	18.3	302	1 GEMSN	osteonectin precu
28	71.5	18.3	303	1 GEHUN	osteonectin precu
29	71.5	18.3	304	1 GEBON	osteonectin precu

30	71.5	18.3	548	2 T16642	hypothetical prote
31	71.5	18.3	1217	1 EGSM5G	epidermal growth f
32	71	18.2	634	1 GERTX1	matrix glycoprotei
33	71	18.2	664	2 S60062	hevin precursor -
34	70.5	18.1	649	2 T00596	hypothetical prote
35	69.5	17.8	300	2 S18992	osteonectin precu
36	69.5	17.8	803	1 IJCH3	integrin, band 3 p
37	68.5	17.6	54	2 A31438	ovomucoid, third d
38	68.5	17.6	1101	2 T16840	hypothetical prote
39	68.5	17.6	1584	2 T22674	hypothetical prote
40	68	17.4	1025	2 T42626	secreted leucine-r
41	67.5	17.3	94	2 C37057	fibronectin recept
42	67.5	17.3	798	2 S01659	integrin beta-1 ch
43	67.5	17.3	799	1 T09059	fibronectin recept
44	67.5	17.3	1964	2 T09059	notch4 - mouse
45	67.5	17.3	1984	2 T13171	probable vitelloge

ALIGNMENTS

RESULT 1
T16859
hypothetical protein T13C2.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T16859
R:Du, Z.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T13C2.
A:Reference number: Z18591
A:Accession: T16859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1296 <DUZ>
A:Cross-references: EMBL:440030; NID:g1055164; PID:g1055165; PIDN:AAA0133.1; CESP:T1
A:Gene: CESP:T13C2.5
A:Introns: 62/3; 93/3; 124/3; 142/1; 212/1; 293/1; 336/2; 406/3; 437/1; 508/1; 584/1;

Query Match 30.8%; Score 120; DB 2; Length 1296;
Best Local Similarity 36.4%; Pred. No. 0.0001;
Matches 20; Conservative 9; Mismatches 24; Indels 2; Gaps 2;

QY 1 CEGFVCPQSGSDCDIM-DGQVPCRCRDCEKESFTCASDGLTYNRCYMDAEC 54
DB 295 CHGKCP-NGOTCOLGVDRRPECKSCSEOTMNSAHVCGTDPGKTYLNECFKLAA 348

RESULT 2
JG0187
follicle-statin-related protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C:Accession: JG0187
R:Okabayashi, K.; Shoji, H.; Onuma, Y.; Nakamura, T.; Nose, K.; Sugino, H.; Asashima, R.
Biochem. Biophys. Res. Commun. 254, 42-48, 1999
A:Title: cDNA cloning and distribution of the Xenopus follicle-statin-related protein.
A:Reference number: JG0187; MUID:99121187
A:Accession: JG0187
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-299 <OKA>

Query Match 28.2%; Score 110; DB 2; Length 299;
Best Local Similarity 31.7%; Pred. No. 0.00033;
Matches 20; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

QY 1 CEGFVCPQSGSDCDIM-DGQVPCRCRDCEKESFTCASDGLTYNRCYMDAECRLGLH 59
DB 28 CANVFC-GAGRECAVTEKGPCTCEKCKSHKRPVCGSGKTYLNNCELHRAACLTGSK 86

OY 60 LHI 62
 Db 87 IQV 89

RESULT 3

S38251
 Follistatin-related protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000
 C:Accession: S38251
 R:Shibanuma, M.; Mashimo, J.; Mita, A.; Kuroki, T.; Nose, K.
 Eur. J. Biochem. 217, 13-19, 1993
 A:Title: Cloning from a mouse osteoblastic cell line of a set of transforming-growth-fac
 A:Reference number: S38251; MUID:94039028
 A:Accession: S38251
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-306 <SRI>
 A:Cross-references: GB:M91360; NID:g349005; PIDN:AAC37633.1; PID:g349006
 C:Superfamily: Kazal proteinase inhibitor homology
 F:48-96/Domain: Kazal proteinase inhibitor homology <KPI5>

Query Match 28.2%; Score 110; DB 2; Length 306;
 Best Local Similarity 31.7%; Pred. No. 0.00033;
 Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;

OY 1 CEGFVCPQGSDDCDIMD-GQPVCRRCRCEKEPSFTCASDGLTYNRCYMDAECALRGH 59
 Db 29 CANVFC-GAGRECAVTEKEGPTCLICIQCKPHKRPVCGSNGKTYLHNCHELRDACLTSK 87

OY 60 LHI 62
 Db 88 IQV 90

RESULT 4

S51361
 Follistatin-related protein precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 17-Nov-2000
 C:Accession: S51361
 R:Zwijnen, A.; Blockx, H.; van Arnhem, W.; Willems, J.; Franssen, L.; Devos, K.; Raymacke
 Eur. J. Biochem. 225, 937-946, 1994
 A:Title: Characterization of a rat C(6) glioma-secreted follistatin-related protein (FRF
 A:Reference number: S51361; MUID:95045570
 A:Accession: S51361
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-306 <2MI>
 A:Cross-references: EMBL:U06864; NID:g536899; PIDN:AAA66063.1; PID:g536900
 C:Superfamily: Kazal proteinase inhibitor homology
 F:48-96/Domain: Kazal proteinase inhibitor homology <KPI5>

Query Match 28.2%; Score 110; DB 2; Length 306;
 Best Local Similarity 31.7%; Pred. No. 0.00033;
 Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;

OY 1 CEGFVCPQGSDDCDIMD-GQPVCRRCRCEKEPSFTCASDGLTYNRCYMDAECALRGH 59
 Db 29 CANVFC-GAGRECAVTEKEGPTCLICIQCKPHKRPVCGSNGKTYLHNCHELRDACLTSK 87

OY 60 LHI 62
 Db 88 IQV 90

RESULT 5

S51362
 Follistatin-related protein - human

C:Species: Homo sapiens (man)
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 17-Nov-2000
 C:Accession: S51362
 R:Zwijnen, A.; Blockx, H.; van Arnhem, W.; Willems, J.; Franssen, L.; Devos, K.; Rayma
 Eur. J. Biochem. 225, 937-946, 1994
 A:Title: Characterization of a rat C(6) glioma-secreted follistatin-related protein (
 A:Reference number: S51361; MUID:95045570
 A:Accession: S51362
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-308 <2MI>
 A:Cross-references: EMBL:U06863; NID:g536897; PIDN:AAA66062.1; PID:g536898
 C:Genetics:
 A:Gene: GDB:FRP
 A:Cross-references: GDB:434868
 C:Superfamily: Kazal proteinase inhibitor homology
 F:50-98/Domain: Kazal proteinase inhibitor homology <KPI5>

Query Match 28.2%; Score 110; DB 2; Length 308;
 Best Local Similarity 31.7%; Pred. No. 0.00034;
 Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;

OY 1 CEGFVCPQGSDDCDIMD-GQPVCRRCRCEKEPSFTCASDGLTYNRCYMDAECALRGH 59
 Db 31 CANVFC-GAGRECAVTEKEGPTCLICIQCKPHKRPVCGSNGKTYLHNCHELRDACLTSK 89

OY 60 LHI 62
 Db 90 IQV 92

RESULT 6

AGRT
 agrin - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C:Accession: JH0399; A38856
 R:Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
 Neuron 6, 811-823, 1991
 A:Title: Structure and expression of a rat agrin.
 A:Reference number: JH0399; MUID:91222570
 A:Accession: JH0399
 A:Molecule type: mRNA
 A:Residues: 1-1779;1799-1959 <RUP>
 A:Cross-references: GB:M64780; NID:g202798; PIDN:AAA0703.1; PID:g202800
 A:Experimental source: embryonic spinal cord
 A>Note: It is uncertain whether Met-1, Met-18, or Met-24 is the initiator
 R:Rupp, F.; Oezcelik, T.; Linal, M.; Peterson, K.; Francke, U.; Scheller, R.
 J. Neurosci. 12, 3535-3544, 1992
 A:Title: Structure and chromosomal localization of the mammalian agrin gene.
 A:Reference number: A38856; MUID:92407628
 A:Accession: A38856
 A:Molecule type: mRNA
 A:Residues: 1780-1798 <RU2>
 A:Cross-references: GB:S44194
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholin
 C:Comment: 90% of rat embryonic transcripts encode the variant labeled below as form
 Y:Choline receptor clustering activity.
 C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G re
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
 F:1-1959/Product: agrin, form 1 #status predicted <AG1>
 F:1-1787;1799-1959/Product: agrin, form 4 #status predicted <AG4>
 F:1-1779;1799-1959/Product: agrin, form 3 #status predicted <AG3>
 F:1-1779;1788-1959/Product: agrin, form 5 #status predicted <AG5>
 F:1-1143;1153-1959/Product: agrin, form 2 #status predicted <AG2>
 F:32-50/Region: hydrophobic
 F:88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
 F:163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
 F:236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
 F:307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
 F:381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
 F:446-494/Domain: Kazal proteinase inhibitor homology <KPI6>

F:511-559/Domain: Kazal proteinase inhibitor homology <KP17>
 F:540-542/Region: motor neuron attachment (L-R-E) motif
 F:596-645/Domain: Kazal proteinase inhibitor homology <KP18>
 F:688-735/Domain: laminin-type EGF-like homology <LE1>
 F:742-786/Domain: laminin-type EGF-like homology <LE2>
 F:814-864/Domain: Kazal proteinase inhibitor homology <KP19>
 F:869-992/Region: serine/threonine-rich
 F:1084-1086/Region: motor neuron attachment (L-R-E) motif
 F:1147-1215/Region: serine/threonine-rich
 F:1224-1257/Domain: EGF homology <EG1>
 F:1287-1442/Domain: laminin G repeat homology <LG1>
 F:1444-1476/Domain: EGF homology <EG2>
 F:1483-1515/Domain: EGF homology <EG3>
 F:1555-1706/Domain: laminin G repeat homology <LG2>
 F:1713-1747/Domain: EGF homology <EG4>
 F:1807-1959/Domain: laminin G repeat homology <LG3>
 F:97-116,105-137,171-191,180-212,244-267,252-284,316-335,324-356,389-408,397-429,454-473
 -1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
 F:145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 25.9%; Score 101; DB 1; Length 1959;
 Best Local Similarity 30.9%; Pred. No. 0.012;
 Matches 21; Conservative 12; Mismatches 31; Indels 4; Gaps 2;

OY 1 CEEFVPCQGGDDINDGQVPCRCRDRCEKPEFTCASDGLTYNNRCYMDAEACLGHL 60
 DB 363 CHGVOC-AREGAVCTVKNRKAECRCQVCSGIDYPCGSDGYTVGSVCESMACITLGRRT 421
 OY 61 HIV--PC 65
 DB 422 QVARRGPC 429

RESULT 7
 AGCH
 agrin precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C:Accession: JH0591; A38857; B38857; I50692
 R:181n, K.W.K.; Ruegg, M.A.; Escher, G.; Kroege, S.; McMahon, U.J.
 Neuron 8, 677-689, 1992
 A:Title: cDNA that encodes active agrin.
 A:Reference number: JH0591; MUID:92322297
 A:Accession: JH0591
 A:Molecule type: mRNA
 A:Residues: 1-1955 <TST>
 A:Cross-references: GB:M94271; NID:9211120; PIDN:AAA48585.1; PID:9211121
 A:Experimental source: brain
 R:Ruegg, M.A.; Tsalm, K.W.K.; Horton, S.E.; Kroege, S.; Escher, G.; Gensch, E.M.; McMahon
 Neuron 8, 691-699, 1992
 A:Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
 A:Reference number: A38857; MUID:92322298
 A:Contents: alternative splicing
 A:Accession: A38857
 A:Molecule type: mRNA
 A:Residues: 1132-1783,1795-1955 <RU2>
 A:Cross-references: GB:M97371
 A:Accession: B38857
 A:Molecule type: mRNA
 A:Residues: 1221-1647,1652-1783,1794-1955 <RU3>
 A:Cross-references: GB:M97372
 A:Note: translation of the nucleotide sequence is not complete
 R:Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
 Dev. Biol. 158, 523-535, 1993
 A:Title: Developmental expression and alternative splicing of chick agrin RNA.
 A:Reference number: I50692; MUID:93345745
 A:Accession: I50692
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1784-1795 <THO>
 A:Cross-references: EMBL:U07271; NID:9459665; PIDN:AAA16788.1; PID:9459666
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r

C:Comment: Alternative splicing produces two inactive proteins: agrin-related protein
 C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G re
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction

F:1-38/Domain: signal sequence #status predicted <SIT>
 F:39-1955/Product: agrin #status predicted <MAT>
 F:39-1783,1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
 F:39-1647,1652-1783,1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
 F:77-126/Domain: Kazal proteinase inhibitor homology <KP11>
 F:152-201/Domain: Kazal proteinase inhibitor homology <KP12>
 F:225-273/Domain: Kazal proteinase inhibitor homology <KP13>
 F:295-344/Domain: Kazal proteinase inhibitor homology <KP14>
 F:370-418/Domain: Kazal proteinase inhibitor homology <KP15>
 F:435-483/Domain: Kazal proteinase inhibitor homology <KP16>
 F:500-548/Domain: Kazal proteinase inhibitor homology <KP17>
 F:584-633/Domain: Kazal proteinase inhibitor homology <KP18>
 F:575-726/Domain: laminin-type EGF-like homology <LE1>
 F:729-773/Domain: laminin-type EGF-like homology <LE2>
 F:801-851/Domain: Kazal proteinase inhibitor homology <KP19>
 F:856-995/Region: serine/threonine-rich
 F:1150-1219/Region: serine/threonine-rich
 F:1233-1264/Domain: EGF homology <EG1>
 F:1294-1448/Domain: laminin G repeat homology <LG1>
 F:1429-1431/Region: motor neuron attachment (L-R-E) motif
 F:1450-1482/Domain: EGF homology <EG2>
 F:1489-1521/Domain: EGF homology <EG3>
 F:1560-1711/Domain: laminin G repeat homology <LG2>
 F:1718-1751/Domain: EGF homology <EG4>
 F:1803-1955/Domain: laminin G repeat homology <LG3>
 F:66-105,94-126,160-180,169-201,233-252,241-273,304-323,312-344,378-397,386-418,443-4
 1482,1489-1500,1494-1510,1512-1521/Disulfide bonds: #status predicted
 F:390,659,764,814/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 25.8%; Score 100.5; DB 1; Length 1955;
 Best Local Similarity 35.6%; Pred. No. 0.014;
 Matches 21; Conservative 11; Mismatches 24; Indels 3; Gaps 2;

OY 10 GSPCDIWDGQVPCRCRDRCEKPEFTCASDGLTYNNRCYMDAEACL-RGLHL-HIYPC 65
 DB 360 GATCVNKRREPCVCCQVCGGRYPVCSNRTYGNCELMNACVLRKRIKHKGPC 418

RESULT 8
 T43060
 agrin - electric ray (Discopyge ommata) (fragment)
 C:Species: Discopyge ommata
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
 C:Accession: T43060
 R:Smith, M.A.; Magill-Solc, G.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; MCM
 submitted to the EMBL Data Library, September 1992
 A:Reference number: Z22308
 A:Accession: T43060
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1328 <SMI>
 A:Cross-references: EMBL:U01423; NID:9213102; PID:9213103; PIDN:AAA49224.1
 C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G re
 C:Keywords: glycoprotein; neuromuscular junction

Query Match 25.6%; Score 100; DB 2; Length 1328;
 Best Local Similarity 35.7%; Pred. No. 0.011;
 Matches 25; Conservative 8; Mismatches 31; Indels 6; Gaps 4;

OY 1 CEEFVPCQGGDDINDGQVPCRCRDR-CEKPEF-TCASDGLTYNNRCYMDAEACLGHL 58
 DB 187 CSLDHC-QVCATCVCSIGRAVCECPSPICPKNQFKVCSDGYTVANECQLKTACQGS 245
 OY 59 HIVIV--PC 65
 DB 246 VINILHOGPC 255

RESULT 9
A53502
folistatin - African clawed frog
N/Alternate names: activin-binding protein
C/Species: *Xenopus laevis* (African clawed frog)
C/Date: 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
C/Accession: A53502; JN0119
R/Hemmati-Brivantou, A.; Kelly, O.G.; Melton, D.A.
Cell 77, 283-295, 1994
A>Title: Folistatin, an antagonist of activin, is expressed in the Spemann organizer at
A/Reference number: A53502; MUID:94221645
A/Accession: A53502
A>Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-319 <HEM>
A/Cross-references: GB:S69801; NID:g546544; PID:AA030638.1; PID:g546545
A/Experimental source: early gastrula, stage 28 head
A/Note: sequence extracted from NCBI Backbone (NCBI:146982, NCBI:146983)
R/Tsushiro, K.; Yamada, R.; Asano, M.; Hashimoto, M.; Muramatsu, M.; Shiohawa, K.
Biochem. Biophys. Res. Commun. 174, 1022-1027, 1991
A>Title: Expression of mRNA for activin-binding protein (folistatin) during early embry
A/Reference number: JN0119; MUID:91128357
A/Accession: JN0119
A>Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 109-267, 'G', 269, 'S', 271-317, 'SYEDTEEEEEEEDPVFISSW' <TAS>
C/Comment: This protein is proposed to act as a regulator of activin in inductive intertra
F/114-164/Domain: kazal proteinase inhibitor homology <KPI>
F/265-316/Domain: kazal proteinase inhibitor homology <KP2>

```

Query Match similarity      23.8%; Score 93; DB 2; Length 319;
Best Local Similarity     36.8%; Pred. No. 0.019;
Matches    25; Conservative   5; Mismatches   30; Indels    8; Gaps    5;

OY      1 CEGFVCPQGGSDCCDIDWD--GQPVCRR-CRDRC--KKEPSFTASDGLTYFNRCYMDAEAC 54
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      245 CEDIGC-SAGKKC-LMDSRVGRGRCALDDLCGEKSDDTVCAISDNTTTPSPCAMKAQAAC 302
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      55 LRGHLHLI 62
          | : | : |
Db      303 STGILLEY 310

RESULT  10
         T46283
hypothetical protein DKFP566D234.1 - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46283
R:Bloecher, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23032
A:Accession: T46283
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-121;122-773 <AAA>
A:Cross-references: EMBL:AL137695
A:Experimental source: fetal kidney; clone DKFP566D234
A>Note: the cDNA sequence contains a -1 frameshift near codon 121
C:Genetics:
;Note: DKFP566D234.1

```

Query Match	23.3%	Score 91:	DB 2:	Length 773;
Best Local Similarity	39.1%;	Pred. No.	0.06;	
Matches 18;	Conservative	5;	Mismatches	23; Indels
				0; Gaps
0y	18	GOPVRCRDRDCEKEPSTFCASDGLTYNRCVMDACLRGLHLHY	63	
Db	9	GQAEACMDLCKRHKPKVPGSGDGEYEHNEHRAACLKOKITTV	54	

```

RESULT 11
S55369
foliistatin - chicken
C.Species: Gallus gallus (chicken)
C.Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 17-Nov-2000
C.Accession: S55369
R.Graham, A.; Lumsden, A.
submitted to the EMBL Data Library, May 1995
A.Description: Interactions between rhombomeres modulate Krox-20 and foliistatin ex
A.Reference number: S55369
A.Accession: S55369
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-343 <GRA>
A.Cross-references: EMBL:X87609; NID:g853833; PINN:CAA60915.1; PID:g853834
C.Superfamily: foliistatin: kazal proteinase inhibitor homology
F.113-163/Domain: kazal proteinase inhibitor homology <KPI>
F.1264-315/Domain: kazal proteinase inhibitor homology <KP2>

```

Query Match	22.6%	Score 88	DB 2	Length 343
Best Local Similarity	35.3%	Pred. No. 0.064		
Matches	24	Conservative	6	Mismatches 30; Indels 8; Gaps 5
QY	1	CEGFVCPGGSGDCLWD---	GGPVCR-GRDRC-EKESFTASDGLTYNNCYDAEAC	54
DB	244	CEIDIC-SAGKKC-LMDKVGGRGRCALCDLCEPSKSDPAVCASDNTYTPSECAKMEAC		301
QY	55	LRGLHLHI	62	
DB	302	SMGVILLEV	309	

RESULT 12
A32141
follistatin 1 precursor - human
N:Contams: follistatin 2 precursor
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1990 #sequence_revision 12-Apr-1996 #text_change 17-Nov-2000
C:Accession: A32141; B32141
R:Shimasaki, S.; Koga, M.; Esch, F.; Cooksey, K.; Mercado, M.; Koba, A.; Deno, N.; Yi
Proc. Natl. Acad. Sci. U.S.A. 85, 4218-4222, 1988
A:Title: Primary structure of the human follistatin precursor and its genomic organiz
A:Reference number: A94199; MUID:88247993
A:Accession: A32141
A:Molecule type: DNA; mRNA
A:Residues: 1-344 <SHT>
A:Cross-references: GB:J03771
A:Experimental source: testis
A:Accession: B32141
A:Molecule type: mRNA
A:Residues: 1-317 <SRNA>
A:Cross-references: GB:M19481; GB:J03771; NID:g182719; PIDN:AAA35851.1; PID:g182721
A:Experimental source: testis
C:Superfamily: follistatin; Kazal proteinase inhibitor homology
C:Keywords: alternative splicing; duplication; gonad
F.1-29/Domain: signal sequence #status predicted <SIG>
F.30-344/Product: follistatin 1 #status predicted <MAT1>
F.30-317/Product: follistatin 2 #status predicted <MAT2>
F.265-316/Domain: kazal proteinase inhibitor homology <KPI6>

Query Match	22.3%	Score 87;	DB 2;	Length 344;
Best Local Similarity	35.7%;	Pred. No. 0.061;		
Matches	25;	Conservative	5;	Mismatches 28; Indels 12; Gaps 5;
QY	1	CEGFVCPGPGSDCDIWD---GQPVC-----CRDRCKEPPSPFCASDGIITYNRCYNDAE	52	
		:		
Db	245	CEDICG-IGKAKC-LMPDKVGRGRSLCEDLCPBKSDEP--VCASDNATTAISCAKMEA	300	
QY	53	ACLRGLHDI 62		

Db 301 ACSSGVLEEV 310

RESULT 13

147079

Follistatin - sheep (fragment)

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 17-Nov-2000

C:Accession: I47079

R:Tisdall, D.J.; Hill, D.; Petersen, G.B.; Fleming, J.S.

J. Mol. Endocrinol. 8, 259-264, 1991

A:Title: Ovine follistatin: characterization of cDNA and expression in sheep ovary during

A:Reference number: I47079; MUID:92337809

C:Accession: I47079

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-337 <TIS>

A:Cross-references: GB:M63123; NID:9165882; PIDN:AAA31522.1; PID:9165883

C:Superfamily: follistatin; Kazal proteinase inhibitor homology <KPI1>

F:107-157/Domain: Kazal proteinase inhibitor homology <KPI1>

F:181-232/Domain: Kazal proteinase inhibitor homology <KPI2>

F:258-309/Domain: Kazal proteinase inhibitor homology <KPI3>

Query Match 22.1%; Score 86; DB 2; Length 337;

Best Local Similarity 35.7%; Pred. No. 0.1;

Matches 25; Conservative 7; Mismatches 26; Indels 12; Gaps 6;

Qy 1 CEGFVCPQGGSDCDIWD---GQPVCR-CRDRC---EKPSFTCASDGLTYVRCYMDAE 52

Db 238 CEDIOC-TGKKC-LMDFVYGRGRCSLCDELCPDSKSEEP--VCASDNATVASECAMKRA 293

Qy 53 ACLRGHLHI 62

Db 294 ACSSGVLEEV 303

RESULT 14

545321

Follistatin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000

C:Accession: S45321; S51666

R:Albano, R.M.; Atkell, R.; Beddington, R.S.P.; Smith, J.C.

Development 120, 803-813, 1994

A:Title: Expression of inhibin subunits and follistatin during postimplantation mouse development.

A:Reference number: S45321; MUID:95324368

C:Accession: S45321

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-343 <ALB>

A:Cross-references: EMBL:229532; NID:9488368; PIDN:CAA82648.1; PID:9488369

R:Ritvos, O.; Tuuri, T.; Ertama, M.; Saltnio, K.; Saxon, L.; Gilbert, S.

submitted to the EMBL Data Library, December 1994

A:Description: Activin disrupts epithelial branching morphogenesis in developing glandu

A:Reference number: S51666

C:Accession: S51666

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 3-240, 'KA', 242-340 <RTV>

A:Cross-references: EMBL:X83377; NID:9603573; PIDN:CAA56291.1; PID:9603574

C:Superfamily: follistatin; Kazal proteinase inhibitor homology

F:114-164/Domain: Kazal proteinase inhibitor homology <KPI>

F:264-315/Domain: Kazal proteinase inhibitor homology <KPI2>

Query Match 22.1%; Score 86; DB 2; Length 343;

Best Local Similarity 35.7%; Pred. No. 0.1;

Matches 25; Conservative 5; Mismatches 28; Indels 12; Gaps 5;

Qy 1 CEGFVCPQGGSDCDIWD---GQPVCR---CRDRCEKPSFTCASDGLTYVRCYMDAE 52

Db 244 CEDIOC-TGKKC-LMDFVYGRGRCSLCDELCPDSKSEEP--VCASDNATVASECAMKRA 299

Qy 53 ACLRGHLHI 62

Db 300 ACSSGVLEEV 309

RESULT 15

A27701

Follistatin precursor - pig

N:Contains: follistatin short form precursor

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 17-Nov-2000

C:Accession: A27701; A40064; B40064; A39969; A47139

R:Shimasaki, S.; Koga, M.; Esch, F.; Mercado, M.; Cooksey, K.; Koba, A.; Ling, N.

Biochem. Biophys. Res. Commun. 152, 717-723, 1988

A:Title: Porcine follistatin gene structure supports two forms of mature follistatin

A:Reference number: A27701; MUID:88209050

C:Accession: A27701

A:Molecule type: DNA

A:Residues: 1-344 <SHI>

A:Cross-references: GB:M19529; NID:9164458; PIDN:AAA31036.1; PID:9164459

R:Esch, F.S.; Shimasaki, S.; Mercado, M.; Cooksey, K.; Ling, N.; Yling, S.; Ueno, N.; Mol. Endocrinol. 1, 849-855, 1987

A:Title: Structural characterization of follistatin: a novel follicle-stimulating hor

A:Reference number: A40064; MUID:91042571

C:Accession: A40064

A:Molecule type: mRNA

A:Residues: 1-344 <ESC1>

A:Cross-references: GB:M36512; GB:M36513; NID:9164461

C:Accession: B40064

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-317 <ESC2>

A:Cross-references: GB:M36512; GB:M36513; NID:9164461

R:Ueno, N.; Ling, N.; Yling, S.Y.; Esch, F.; Shimasaki, S.; Guillemin, R.

Proc. Natl. Acad. Sci. U.S.A. 84, 8282-8286, 1987

A:Title: Isolation and partial characterization of follistatin: a single-chain M-r 35

A:Reference number: A39969; MUID:88068578

C:Accession: A39969

A:Molecule type: protein

A:Residues: 30-37 <UEN>

R:Sugino, K.; Kurosawa, N.; Nakamura, T.; Takio, K.; Shimasaki, S.; Ling, N.; Titani, J. Biol. Chem. 268, 15579-15587, 1993

A:Title: Molecular heterogeneity of follistatin, an activin-binding protein. Higher a

A:Reference number: A47139; MUID:93340158

C:Accession: A47139

A:Status: preliminary

A:Molecule type: protein

A:Residues: 30-35, 120-128, 284-292, 312-344 <SUC>

C:Comment: Follistatin inhibits release of pituitary follicle stimulating hormone.

C:Genetics:

A:Introns: 29/1; 93/1; 166/1; 241/1; 318/1

C:Superfamily: follistatin; Kazal proteinase inhibitor homology

C:Keywords: alternative splicing; glycoprotein; monomer

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-344/Product: follistatin #status predicted <MAT>

F:30-317/Product: follistatin, short splice form #status predicted <MAS>

F:114-164/Domain: Kazal proteinase inhibitor homology <KPI8>

Query Match 22.1%; Score 86; DB 1; Length 344;

Best Local Similarity 35.7%; Pred. No. 0.1;

Matches 25; Conservative 7; Mismatches 26; Indels 12; Gaps 6;

Qy 1 CEGFVCPQGGSDCDIWD---GQPVCR-CRDRC---EKPSFTCASDGLTYVRCYMDAE 52

Db 245 CEDIOC-TGKKC-LMDFVYGRGRCSLCDELCPDSKSEEP--VCASDNATVASECAMKRA 300

Qy 53 ACLRGHLHI 62

Db 301 ACSSGVLEEV 310

Wed Feb 27 08:23:18 2002

us-09-819-136-2_copy_93_157.rpr

Page 6

Search completed: February 26, 2002, 01:29:34
Job time: 413 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:02 ; Search time 69.26 Seconds
(without alignments)
21.119 Million cell updates/sec

Title: US-09-819-136-2_COPY_93_157

Perfect score: 390
Sequence: 1 CEGFVCPQGSDCDIMDGP.....RCYMDADACLRGLHLITVPC 65

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	25.9	1940	2	US-08-644-271-30
2	95	24.4	71	2	US-08-972-008-4
3	95	24.4	263	2	US-08-972-008-2
4	84	21.5	380	3	US-08-468-846-2
5	84	21.5	380	4	US-08-915-096A-2
6	80.5	20.6	374	2	US-08-820-170A-25
7	80.5	20.6	374	4	US-09-055-699-25
8	80.5	20.6	374	4	US-09-273-565-25
9	80.5	20.6	374	4	PCR-US95-06385-2
10	71.5	18.3	303	6	5340934-13
11	71	18.2	664	1	US-08-421-661-6
12	70.3	18.1	317	4	US-09-383-586-20
13	68	17.4	771	4	US-09-188-930-183
14	67.5	17.3	94	4	US-07-728-215-37
15	67	17.2	184	1	US-08-211-942-7
16	67	17.2	186	1	US-08-211-942-9
17	67	17.2	1525	3	US-09-191-647-2
18	67	17.2	1525	4	US-09-540-245A-2
19	67	17.2	1525	3	US-09-540-153-2
20	66.5	17.1	160	3	US-09-191-647-5
21	66.5	17.1	160	4	US-09-540-245A-5
22	66.5	17.1	160	4	US-09-540-153-5
23	66	16.9	578	4	US-08-981-392-13
24	65.5	16.8	94	2	US-07-728-215-35
25	65.5	16.8	798	2	US-07-728-215-30
26	65	16.7	104	3	US-09-191-647-13
27	65	16.7	104	4	US-09-540-245A-13

28	65	16.7	104	4	US-09-540-153-13	Sequence 13, Appl
29	65	16.7	480	3	US-08-923-454A-18	Sequence 18, Appl
30	64.5	16.5	1139	1	US-08-537-210A-4	Sequence 4, Appl
31	64.5	16.5	1139	4	US-09-113-825-4	Sequence 19, Appl
32	64.5	16.5	2703	1	US-08-185-432-19	Sequence 3, Appl
33	63.5	16.3	103	1	US-08-211-942-3	Sequence 17, Appl
34	63.5	16.3	368	1	US-08-211-942-17	Sequence 5, Appl
35	63	16.2	122	2	US-08-232-087A-7	Sequence 1, Appl
36	62.5	16.0	73	2	US-08-972-008-5	Sequence 3, Appl
37	62.5	16.0	86	2	US-08-744-670-1	Sequence 2, Appl
38	62.5	16.0	86	2	US-09-149-933-1	Sequence 2, Appl
39	62.5	16.0	86	4	US-09-065-019-3	Sequence 2, Appl
40	62.5	16.0	595	1	US-08-225-989-2	Sequence 2, Appl
41	62.5	16.0	595	1	US-08-570-923-2	Sequence 2, Appl
42	62.5	16.0	595	1	US-08-580-014-2	Sequence 2, Appl
43	62.5	16.0	595	2	US-08-232-087A-2	Sequence 2, Appl
44	62.5	16.0	595	4	US-09-079-785-2	Sequence 2, Appl
45	62.5	16.0	595	4	US-09-006-353A-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; NUMBER OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rat Agt1n
; LOCATION: 1..1940
; OTHER INFORMATION:
; US-08-644-271-30

Query Match 25.9%; Score 101; DB 2; Length 1940;
Best Local Similarity 30.9%; Pred. No. 0.0034;
Matches 21; Conservative 12; Mismatches 31; Indels 4; Gaps 2;

QY 1 CEGFVCPQGSDDIMDGPVRCRDRCKEPPS--FTCASDGLTYNRCYMDAFCARG 60
DB 363 CHGVOC-AFGAVCTVKNRKAECORVCSGIYDPVCGSDGVYTGSCVCELSNACTLGREI 421
QY 61 HIV---PC 65
DB 422 QVARRGPC 429

RESULT 2
US-08-972-008-4
; Sequence 4, Application US/08972008
; Patent No. 5942420
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: No. 5942420e1 Molecules of the Follistatin-Related
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,008
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-972-008-4

Query Match 24.4%; Score 95; DB 2; Length 71;
Best Local Similarity 37.3%; Pred. No. 0.00046;
Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

QY 1 CEGFVCPQGSDDIMDGPVRCRDRCKEPPS--FTCASDGLTYNRCYMDAFCARG 57
DB 3 CDGVEG-GPGKACRMIGRPRCECAPDCSGIPARLVCGSDGATYRDECELRARC-RG 59

RESULT 3
US-08-972-008-2
; Sequence 2, Application US/08972008
; Patent No. 5942420
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: No. 5942420e1 Molecules of the Follistatin-Related
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,008
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: MNI-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-972-008-2

Query Match 24.4%; Score 95; DB 2; Length 263;
Best Local Similarity 37.3%; Pred. No. 0.0019;
Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

QY 1 CEGFVCPQGSDDIMDGPVRCRDRCKEPPS--FTCASDGLTYNRCYMDAFCARG 57
DB 99 CDGVEG-GPGKACRMIGRPRCECAPDCSGIPARLVCGSDGATYRDECELRARC-RG 155

RESULT 4
US-08-468-846-2
; Sequence 2, Application US/08468846
; Patent No. 6074839
; GENERAL INFORMATION:
; APPLICANT: Meissner, Paul
; APPLICANT: Fuldner, Rebecca
; APPLICANT: Fei-Wel, Ying
; APPLICANT: Adams, Mark
; TITLE OF INVENTION: TRANSFORMING GROWTH FACTOR ALPHA HI
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; ADDRESSEE: STUART & OLSTEIN
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,846
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/208,008
; FILING DATE: 08-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 325800-465
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-846-2

Query Match 21.5%; Score 84; DB 3; Length 380;
Best Local Similarity 30.3%; Pred. No. 0.047;
Matches 20; Conservative 9; Mismatches 27; Indels 10; Gaps 3;

OY 3 GFVCPQSGSDCDIMDGPVCRRCRCEKEPFTCASDGLTYYNRCYMDAACLRLGLHLH 62
DB 85 GGVCKEDG-----DGLK-CACQFCQHTNYIPVCGSNGDTYQNECFLLRRAACKHQKEITV 137
OY 63 V---PC 65
DB 138 IARGPC 143

RESULT 5
US-08-915-096A-2
Sequence 2, Application US/08915096A
Patent No. 6265543
GENERAL INFORMATION:
APPLICANT: Melasner, Paul S.
APPLICANT: Fuldner, Rebecca A.
TITLE OF INVENTION: Transforming Growth Factor Alpha HI
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915.096A
FILING DATE: 20-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/468, 846
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/208, 008
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PFI10D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-096A-2

Query Match 21.5%; Score 84; DB 4; Length 380;
Best Local Similarity 30.3%; Pred. No. 0.047;
Matches 20; Conservative 9; Mismatches 27; Indels 10; Gaps 3;

OY 3 GFVCPQSGSDCDIMDGPVCRRCRCEKEPFTCASDGLTYYNRCYMDAACLRLGLHLH 62
DB 85 GGVCKEDG-----DGLK-CACQFCQHTNYIPVCGSNGDTYQNECFLLRRAACKHQKEITV 137
OY 63 V---PC 65
DB 138 IARGPC 143

RESULT 6
US-08-820-170A-25
Sequence 25, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIMARA
APPLICANT: Takeshi, MATANABE
APPLICANT: Masato, HORIE
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820.170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-820-170A-25

Query Match 20.6%; Score 80.5; DB 2; Length 374;
Best Local Similarity 28.6%; Pred. No. 0.11;
Matches 18; Conservative 10; Mismatches 34; Indels 1; Gaps 1;

OY 1 CEGFVCPQSGSDCDIMDGPVCRRCRCEKEPFTCASDGLTYYNRCYMDAACLRLGLHLH 60
DB 69 CCFNCKPGEDECLRIQD-IVTCVQCFKCNNDYIPVCGSNGESYQNECFLLRRAACKHQKEITV 127
OY 61 HIV 63
DB 128 LVV 130

RESULT 7
US-09-055-699-25
Sequence 25, Application US/09055699
Patent No. 6005088
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIMARA

```

; SEQ ID NO 25
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-273-565-25

Query Match          20.6%; Score 80.5; DB 4; Length 374;
Best Local Similarity 28.6%; Pred.No.0.11;
Matches 18; Conservative 10; Mismatches 34; Indels 1; Gaps 1

QY      1 CEGFVCPQGSDDIDMDGQVPCRCRRCCEKPEPFTCASDGLTYNNRCYMDAEACLRGLHL 60
       | : - - - - - | : - - - - - | : - - - - - | : - - - - - | : - - - - -
Db      69 CDNTNCKDEBCLIRIGD-TVTVCVCFKCNNDIYPVCGSGNSGYONECYTLROACKKOQSEI 127

QY      61 HIV 63
       |
Db      128 LVV 130

RESULT    9
PCT-US95-06385-2
; Sequence 2, Application PC/TUS9506385
; GENERAL INFORMATION:
; APPLICANT: WEI, ET AL.
; TITLE OF INVENTION: Transforming Growth Factor Alpha III
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06385
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NONE
; FILING DATE: NONE
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.,
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-351
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
PCT-US95-06385-2

Query Match          20.6%; Score 80.5; DB 5; Length 374;
Best Local Similarity 28.6%; Pred.No.0.11;
Matches 18; Conservative 10; Mismatches 34; Indels 1; Gaps 1;

QY      1 CEGFYCPOGGSDCDIDWGQPYVCRRCRCKEPPSFTCASDGLTYNNRCYMAEAACLRLGIHL 60
       | : - - - - - | : - - - - - | : - - - - - | : - - - - - | : - - - - -
Db      69 CDNTNCKDEBCLIRIGD-TVTVCVCFKCNNDIYPVCGSGNSGYONECYTLROACKKOQSEI 127

```


DB 128 LVV 130

RESULT 10

5340934-13

PATENT NO. 5340934

APPLICANT: TERMINE, JOHN D.; YOUNG, MARIAN F.; FISHER, LARRY W.

ROBEY, PAMELA G.

TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS

NUMBER OF SEQUENCES: 13

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/432,044

FILING DATE: 03-NOV-1989

SEQ ID NO:13:

LENGTH: 303

5340934-13

Query Match

Best Local Similarity 18.3%; Score 71.5; DB 6; Length 303;

Matches 24; Conservative 13; Mismatches 27; Indels 15; Gaps 7;

DB 72 CQNHHC-KKHCVCCELDENNTPMKVCODPTSCAPRIGEFKVCSDNKKTFDSSCHFPATKC 130

OY 1 CEGFVCPQGGSDCDIWDGQ-PVCRCD--RCEK---EPSEFTCASDGLTYNRCYMDAEAC 54

DB 72 CQNHHC-KKHCVCCELDENNTPMKVCODPTSCAPRIGEFKVCSDNKKTFDSSCHFPATKC 130

OY 55 -LRG-----LHL-HIVPC 65

DB 131 TLEGTKKGHKLHLDYIGPC 149

RESULT 11

US-08-421-661-6

Sequence 6, Application US/08421661

Patent No. 5807993

GENERAL INFORMATION:

APPLICANT: French, Cynthia K

APPLICANT: Yamamoto, Karen K

APPLICANT: Chow, Phoebe M

APPLICANT: Alido, Nemesias T

TITLE OF INVENTION: Antigen Related to Inflammatory Diseases

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: David J. Oldenkamp

STREET: 2029 Century Park East, Suite 3800

CITY: Los Angeles

STATE: California

COUNTRY: U.S.

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/421,661

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oldenkamp, David J

REGISTRATION NUMBER: 29,421

REFERENCE/DOCKET NUMBER: 109-067

TELECOMMUNICATION INFORMATION:

TELEPHONE: 310-788-5000

TELEFAX: 310-277-1297

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 664 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-421-661-6

Query Match

Best Local Similarity 18.2%; Score 71; DB 1; Length 664;

Matches 18; Conservative 9; Mismatches 26; Indels 6; Gaps 3;

OY 1 CEGFVCPQGGSDCDI-WDGPVCRCDRCRCEKPS----FTCASDGLTYNRCYMDAEAC 54

DB 433 CMSFQC-KRCHICKADQGGPHRCVCODPVYCPPTFKPLDQVCGTIDNOTYASSCHLPATKC 490

RESULT 12

US-09-383-586-20

Sequence 20, Application US/09383586

Patent No. 6242419

GENERAL INFORMATION:

APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew

APPLICANT: Abernethy, Nevin

APPLICANT: Onrust, Rene

APPLICANT: Kumbale, Anand

APPLICANT: Murison, Greg

TITLE OF INVENTION: Compounds isolated from stromal cells

FILE REFERENCE: 11000.1037c1

CURRENT APPLICATION NUMBER: US/09/383,586

CURRENT FILING DATE: 1999-08-26

NUMBER OF SEQ ID NOS: 38

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 20

LENGTH: 317

TYPE: PRT

ORGANISM: Mouse

US-09-383-586-20

Query Match

Best Local Similarity 18.1%; Score 70.5; DB 4; Length 317;

Matches 20; Conservative 5; Mismatches 30; Indels 5; Gaps 4;

OY 1 CEGFVCPQGGSDC-DIWDGQPVCRCDRCRCEKPS--FTCASDGLTYNRCYMDAEACLRG 57

DB 97 CDGVEEC-GPEKACRNAGASNNCECVNCEGPFAGFQVCGSDATYRDECELRARC-RG 154

RESULT 13

US-09-188-930-183

Sequence 183, Application US/09188930A

Patent No. 6150502

GENERAL INFORMATION:

APPLICANT: Watson, James D.

APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew

APPLICANT: Onrust, Rene

APPLICANT: Murison, James Greg

TITLE OF INVENTION: Compositions Isolated From Skin Cells

FILE REFERENCE: 11000.1011c1

CURRENT APPLICATION NUMBER: US/09/188,930A

CURRENT FILING DATE: 1998-11-09

NUMBER OF SEQ ID NOS: 348

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 183

LENGTH: 771

TYPE: PRT

ORGANISM: Rat

FEATURE:

NAME/KEY: UNSURE

LOCATION: (717)...(717)

US-09-188-930-183

Query Match

17.4%; Score 68; DB 4; Length 771;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:39 ; Search time 162.6 Seconds
(without alignments)
29.611 Million cell updates/sec

Title: US-09-819-136-2_COPY_93_157

Perfect score: 390

Sequence: 1 CEGFVCGQGSDDCDIMDGP.....RCYMDAECACGLHLATVPC 65

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	327	83.8	216	22	AAB88333 Human membrane or
2	110	28.2	308	18	AAW17860 Follistatin relate
3	98	23.1	233	20	AAV14569 Rat apoptosis inh
4	98	25.1	256	20	AAV14570 Rat apoptosis inh
5	97	24.9	815	21	AAB19728 Human SECX clone A
6	97	24.9	842	21	AAB19727 Human SECX clone A
7	97	24.9	842	22	AAB82300 Human follistatin-
8	95	24.4	71	20	AAV06202 Human FCMF follist
9	95	24.4	263	20	AAV14571 Human apoptosis in
10	95	24.4	263	20	AAV06201 Follistatin module
11	95	24.4	263	20	AAV01098 Human follistatin-

12	95	24.4	263	22	AAU12413 Human PRO1308 poly
13	95	24.4	263	22	AAV72875 Human PRO1308 prot
14	95	24.4	263	22	AAV72629 Human follistatin-
15	95	24.4	263	22	AAV31190 Amino acid sequenc
16	95	24.4	263	22	AAV50987 Human PRO1308 prot
17	93	23.8	319	16	AAV74600 Follistatin. Synt
18	91.5	23.5	413	22	AAV93754 Human protein sequ
19	91	22.3	343	21	AAB27655 Human PRO2
20	91	22.3	343	21	AAB33458 Human PRO205 prote
21	91	22.3	983	21	AAV70775 Human EDP-binding
22	87	22.3	304	13	AAV20063 Human EDP-binding
23	87	22.3	317	10	AAV93396 Human follistatin-
24	87	22.3	317	22	AAV72630 Human polypeptide
25	87	22.3	344	22	AAV33389 Human polypeptide
26	87	22.3	345	22	AAV70351 Human follistatin
27	87	22.3	432	22	AAV41175 Human polypeptide
28	86	22.1	304	13	AAV20061 Pig EDP-binding pr
29	86	22.1	304	13	AAV20062 Rat EDP-binding pr
30	84	21.5	280	18	AAV09406 Transforming growt
31	83.5	21.4	257	21	AAV76102 Murine skin cell p
32	83.5	21.4	257	22	AAV56041 Skin cell protein
33	83	21.3	195	21	AAV23270 Balanus amphitrite
34	82.5	21.2	433	20	AAV02602 HPTG protein of h
35	82.5	21.2	436	19	AAV74722 Human secreted pro
36	82.5	21.2	436	20	AAV41753 Human PRO771 prote
37	82.5	21.2	436	21	AAV44309 Human PRO771 (UNQ4
38	82.5	21.2	436	21	AAV24414 Human PRO771 prote
39	82.5	21.2	436	22	AAV23906 Human EST encoded
40	82.5	21.2	436	22	AAV24064 Human EST encoded
41	82.5	21.2	436	22	AAU12373 Human PRO771 poly
42	82.5	21.2	447	22	AAE03313 Human gene 2 encod
43	80.5	20.6	347	22	AAV01225 Human prostate spe
44	80.5	20.6	374	18	AAV37497 Human TWP-2. Homo
45	80.5	20.6	374	18	AAV07663 Human transforming

ALIGNMENTS

RESULT 1	
AAV88333	
ID AAB88333	standard; Protein; 216 AA.
XX	
AC AAB88333;	
XX	
DT 23-MAY-2001	(first entry)
XX	
DE Human membrane or secretory protein clone PSEC0040.	
XX	
KW Human; secretory protein; membrane protein; vaccine; gene therapy;	
KW Rheumatoid arthritis; diabetes.	
XX	
OS Homo sapiens.	
XX	
PN EPI067182-A2.	
XX	
PD 10-JAN-2001.	
XX	
PF 07-JUL-2000; 2000EP-0114090.	
XX	
PR 08-JUL-1999; 99UP-0194179.	
PR 11-JAN-2000; 2000UP-0118775.	
PR 02-MAY-2000; 2000UP-0183766.	
XX	
PA (HELI-) HELIX RES INST.	
XX	
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;	
XX	
DR WPI: 2001-093989/11.	
XX	
PT N-PSDB: AAF93760.	
XX	
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -	

XX Claim 1: SEQ ID 34: 609pp + CD ROM, English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 XX Sequence 216 AA;
 XX

Query Match 83.8%; Score 327; DB 22; Length 216;
 Best Local Similarity 100.0%; Pred. No. 1.2e-26;
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CEGFVCPQGSDDCDIWDGQPVCRRCRCEKPEPFTCASDGLTYNRCYMDAEC 54
 Db 93 cegfvcpqgsddcdiwdgqpvcrrcrcekepeftcasdgltyncymdaec 146

RESULT 2
 AAM17860
 ID AAM17860 standard; Protein; 308 AA.
 AC AAM17860;
 XX
 DT 04-FEB-1998 (first entry)
 XX
 DE Follistatin related protein.
 XX
 KW Rheumatoid arthritis; auto-antigen; FRP; diagnosis;
 KW prediction; synovial cell; Follistatin related protein.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 175 /note="encoded by AT"
 XX
 PN WO9717441-A1.
 PD 15-MAY-1997.
 XX
 PF 06-NOV-1996; 96WO-JP03250.
 XX
 PR 07-NOV-1995; 95JP-0288957.
 XX
 PA (KANF) KANEKA CORP.
 XX
 PI Kishimura M, Nakao K, Osakada F, Osaki S, Tanaka M;
 XX WPI, 1997-281030/25.
 DR N-PSDB; AAT68831.
 XX

XX Auto-antigen from synovial cells of rheumatoid arthritis patients -
 PT binds to antibodies present in these patients, for diagnosis and
 PT prediction of the disease
 XX
 PS Claim 6: Pages 38-40; 61pp; Japanese.
 XX
 CC The present sequence is the rheumatoid arthritis (RA)
 CC auto-antigen follistatin related protein (FRP), which can be used
 CC to diagnose and predict the development of RA by reaction with
 CC antibodies in biological specimens, e.g. sera, from patients.
 CC RNA was isolated from synovial cells from a RA patient and used to
 CC construct a cDNA library. This was screened using IgG separated
 CC from the synovial fluid of a RA patient. Active clones were
 CC isolated in a cloning vector, and inserted into an expression
 CC vector for the transformation of E. coli NM522. Transformants on
 CC culture express clone A peptide and FRP into the culture medium.
 CC
 XX Sequence 308 AA;
 XX

Query Match 28.2%; Score 110; DB 18; Length 308;
 Best Local Similarity 31.7%; Pred. No. 0.00063;
 Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;
 QY 1 CEGFVCPQGSDDCDIWDGQPVCRRCRCEKPEPFTCASDGLTYNRCYMDAEC 59
 Db 31 canvfc-gagrecavtekepeftcicqkphkrpvcgsmgktylnhceihndactlgtgsk 89
 QY 60 LHI 62
 Db 90 lqv 92

RESULT 3
 AAY14569
 ID AAY14569 standard; Protein; 233 AA.
 AC AAY14569;
 XX
 DT 14-SEP-1999 (first entry)
 XX
 DE Rat apoptosis inhibitory factor PCRF35 mature protein.
 XX
 KW Neotrophic; neuroprotective; apoptosis inhibitory factor; PCRF35; rat;
 KW pheochromocytoma cell; PC12 cell; human; bcl12; nerve cell death;
 KW Parkinson's disease; Alzheimer's disease.
 XX
 OS Rattus sp.
 XX
 PN WO9931237-A1.
 PD 24-JUN-1999.
 XX
 PF 11-DEC-1998; 98WO-JP05609.
 XX
 PR 14-MAY-1998; 98JP-0131634.
 PR 12-DEC-1997; 97JP-0343112.
 XX
 PA (TAIS) TAISHO PHARM CO LTD.
 XX
 PI Ohsawa Y, Uchiyama Y;
 XX WPI, 1999-395180/33.
 DR N-PSDB; AAX79359.
 XX
 PT New protein obtained by culturing pheochromocytoma transformed by a
 PT bcl2 gene as a protooncogene, useful in remedies for
 PT apoptosis-caused diseases. e.g. Alzheimer disease
 XX
 PS Claim 3: Page 38-39; 50pp; Japanese.
 XX
 CC This sequence represents the mature apoptosis inhibitory factor PCRF35.

RESULT	5
AAB19728	
ID	AAB19728 standard; Protein; 815 AA.
XX	
AC	AAB19728;
XX	

DE	Human SECC Clone AC012614_1.0.0.123 encoded protein.
XX	
KW	SECC; human; diagnosis; therapy; cell adhesion;
KW	neurodegenerative disease; epilepsy; tissue regeneration
XX	
OS	Homo sapiens.
XX	

PD	19-OCT-2000.
XX	
PF	07-APR-2000; 2000WO-US09392.

PR 09-APR-1999; 990US-0128514.
PR 03-MAR-2000; 2000US-0128514.
XX
PA
(CURA-) CURAGEN CORP.
XX

XX
DR WPI, 2000-679487/66.
DR N-PSDB; AAA88797.
DR

PT SECX polypeptides and the nucleic acids that encode them, useful for
PT diagnosing, preventing and treating e.g. cancers, inflammation,
PT arthritis and immunological disorders -

PS Claim 5; Fig 9; 143pp; English
XX
CC The present sequence is that of

membrane-associated proteins encoded by them (see AAB19720-34).

CC the detection, diagnosis and treatment (including gene therapy) of
CC a broad range of pathological states. Clone AC012614.1, 0.123 has
CC similarity to cell adhesion molecules, follistatin, roundabout and
CC frazzled, which are involved in neuronal development and

CC treatment of nerve trauma, neurodegenerative disorders, epilepsy,
CC mental health conditions, tissue regeneration in vivo and in
CC vitro, and female reproductive system disorders and pregnancy.

SQ **Sequence** **815 AA;**

Query Match 24.9%; Score 97; DB 21; Length 815;
Best Local Similarity 35.6%; Pred. No. 0.039;
Matches 21; Conservative 10; Mismatches 22; Indels 6; Gaps 3

[illegible]

RESULT 6
AAB19727

XX AAB19727;
AC
XX

DE Human SECX Clone 4324229-2 encoded protein.
 XX SECX; human; diagnosis; therapy; surface adhesion protein;
 KW antitumour; neurological disorder; developmental disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200061754-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000WO-US09392.
 XX
 PR 09-APR-1999; 99US-0128514.
 XX
 PR 03-MAR-2000; 2000US-0128514.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Fernandez E, Vernet C, Shinkets R;
 XX
 DR WPI; 2000-679487/66.
 DR N-PSDB; AAA88796.
 XX
 PT SECX polypeptides and the nucleic acids that encode them, useful for
 PT diagnosing, preventing and treating e.g. cancers, inflammation,
 PT arthritis and immunological disorders -
 XX
 PS Claim 1; Fig 8; 143pp; English.
 XX
 CC The present sequence is that of surface adhesion protein-like
 CC variant encoded by SECX Clone 4324229-2 (see AAA88796). High
 CC expression was detected in the lung. The invention provides
 CC novel SECX polynucleotides (see AAA88789-904) and the secreted or
 CC membrane-associated proteins encoded by them (see AAB19720-34).
 CC SECX polynucleotides, polypeptides and antibodies can be used in
 CC the detection, diagnosis and treatment (including gene therapy) of
 CC a broad range of pathological states. Therapeutic indications for
 CC targeting 4324299 include selected lung, breast and ovarian
 CC carcinomas. 4324299 has similarity to human limbic system
 CC associated membrane protein (LAMP) and may therefore be important
 CC in nerve growth and differentiation, epilepsy, Alzheimer's
 CC disease and schizophrenia. It also shows similarity to portions
 CC of human Down syndrome-cell adhesion molecule (DS-CAM2), and may
 CC therefore be useful in the detection, diagnosis and therapy of
 CC developmental and neurological abnormalities such as Down
 CC syndrome, mental retardation, holoprosencephaly, agenesis of the
 CC corpus callosum and schizencephaly.
 CC
 XX Sequence 842 AA;
 SQ
 OY 9 QGSDCDI-WDGPVRCRDRCKEKEPSF--TCASDGLTYVYRCYMDAECRLGHLHIV 63
 Db 72 rgsrvlsirkgepcqcleac-rpsvpcgsdgfyfhencklhraacldgrricvi 128
 Matches 21; Conservative 10; Mismatches 22; Indels 6; Gaps 3;
 Query Match 24.9%; Score 97; DB 21; Length 842;
 Best Local Similarity 35.6%; Pred. No. 0.04;
 RESULT 7
 AAB82300
 ID AAB82300 standard; Protein; 842 AA.
 AC AAB82300;
 XX
 XX 09-JUL-2001 (first entry)
 DE Human follistatin-related protein zfst4.
 XX
 XX Follistatin-related protein; zfst4; human; antiinflammatory;
 KW neuroprotective; gynaecological; antiarteriosclerotic;
 KW dermatological; antiallergic; antibacterial; immunosuppressive;

KW antiulcer; vulnereary; ophthalmological; antiporiatic;
 KW antiasthmatic; antiout; analgesic; osteopathic; contraceptive;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Peptide
 FT 1..22
 FT /label= Signal_peptide
 FT 23..842
 FT /label= Mature_protein
 FT 23..64
 FT /note= "hydrophilic short linker domain"
 FT 65..133
 FT /note= "follistatin homology domain"
 FT 134..173
 FT /note= "alpha-helical linker region"
 FT 177..250
 FT /note= "calmodulin domain"
 FT 251..334
 FT /note= "I-set Ig domain 1"
 FT 335..432
 FT /note= "I-set Ig domain 2"
 FT 433..842
 FT /note= "C-terminal domain"
 XX
 PN WO200132871-A2.
 XX
 PD 10-MAY-2001.
 XX
 PF 31-OCT-2000; 2000WO-US30013.
 XX
 PR 03-NOV-1999; 99US-0432359.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Conklin DC, Sheppard PO;
 XX
 DR WPI; 2001-328792/34.
 DR N-PSDB; AAF30913, AAF30917.
 XX
 PT New follistatin-related protein, zfst4, for treating central nervous
 PT system, reproductive, haematopoietic and bone-related disorders -
 XX
 PS Claim 1; Page 82-85; 91pp; English.
 XX
 CC The present sequence is that of a novel human follistatin-related
 CC protein zfst4. The sequence is deduced from an isolated cDNA
 CC clone (see AAF30914). Polypeptides comprising amino acid residues
 CC 23-64, 65-133, 134-173, 65-250, 65-334, 65-432, 65-842,
 CC 134-250, 134-334 and 134-432 of this sequence are claimed. zfst4
 CC is likely to play a role in development and differentiation,
 CC pathogenesis of atherosclerosis, regulation of the
 CC gonadal-pituitary-hypothalamic axis, tooth and bone formation,
 CC regulation of gonadal hormone production, spermatogenesis,
 CC hypothalamic oxytocin secretion, proliferation and differentiation of
 CC erythroid progenitors, haematopoiesis, host defence and neuron
 CC survival. Expression is largely confined to brain, kidney, pancreas,
 CC adrenal gland, prostate and testis. zfst4 polypeptides,
 CC polynucleotides, agonists and antagonists are useful for inhibiting
 CC inflammatory response, stimulating a reduction in number and activity
 CC of inflammatory cells, diminishing oedema and inflammation, treating
 CC acute inflammation conditions, bursitis, chronic inflammation
 CC demyelinating polyneuropathy, contact dermatitis, contact
 CC vulvovaginitis, myositis, sepsis, ulcerative colitis, acute
 CC renal failure, pancreatitis and neonatal bronchopulmonary dysplasia,
 CC ocular injuries such as corneal injury from burns or penetration of
 CC a foreign body or ocular inflammatory diseases such as uveitis, to
 CC alleviate chronic itching and inflammation associated with
 CC dermatological conditions and skin diseases such as eczema,
 CC neurodermatitis, allergy, psoriasis, xerosis, insect bites and
 CC burns, symptoms associated with gout, asthma, carpal tunnel
 CC syndrome, systemic lupus erythematosus, multiple sclerosis and

CC myasthenia gravis. They are also useful for neutralizing the
CC activities of transforming growth factor-beta family members,
CC treating stenosis after angioplasty, atherosclerosis, treating
CC brain and spinal cord injuries, inflammatory disorders, for
CC modulating haematopoiesis and differentiation of erythroid
CC progenitors, to stimulate spermatogenesis, for treating male and
CC female reproductive disorders, in repair of congenital,
CC trauma-induced or surgical resection of bone, in cosmetic surgery,
CC for treating cartilage defects or disorders and to stimulate wound
CC healing, tissue repair and for tooth and bone formation. Zfsta4
CC antagonists are useful as contraceptives.
CC
XX
SQ Sequence 842 AA:

Query Match 24.9%; Score 97; DB 22; Length 842;
Best Local Similarity 35.6%; Pred. No. 0.04;
Matches 21; Conservative 10; Mismatches 22; Indels 6; Gaps 3;

OY 9 CGSDCDI--WDGPPVCRDRCEKEPSF--TCASDGLTYNNRCYMDAEACLRGHTIV 63
Db 72 rgsrvlsrkgepcqcleac--rpsvypvcsgdgrfyencklnraacilgkrltvl 128

RESULT 8

ID AAY06202 standard; Protein: 71 AA.

AAV06202;

16-AUG-1999 (first entry)

Human FCMF follistatin cysteine-rich domain 1.

KW Follistatin module containing protein; FCMF; human; cell survival;
KW cell differentiation; cell proliferation; cancer; preclampsia;
therapy; TANGO 91.

OS Homo sapiens.

PN WO9925371-A1.

27-MAY-1999.

17-NOV-1998; 98WO-US24535.

17-NOV-1997; 97US-0972008.

(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

PI Holtzman DA;

WPI: 1999-337874/28.

Nucleic acid encoding follistatin-module containing protein

Claim 31; Page 86; 92pp; English.

CC The present sequence represents follistatin cysteine-rich domain 1
CC of human follistatin-module containing protein FCMF (see also
CC AAY06201). The 2 follistatin cysteine-rich domains of FCMF are 55%
CC homologous to the follistatin domains of human follistatin.
CC Proteins of the novel FCMF contain at least one follistatin
CC cysteine-rich domain. The invention provides FCMF nucleic acids
CC and proteins, recombinant expression vectors, host cells and
CC transgenic animals. FCMF polypeptides are used: (i) to screen for
CC modulators of FCMF activity; (ii) to raise antibodies, used to
CC detect FCMF in immunoassays, also as therapeutic modulators and for
CC affinity purification of FCMF; or (iii) as modulators. Agents that
CC modulate activity of FCMF, or expression of FCMF-encoding nucleic
CC acid, are useful for treating diseases characterized by aberrant
CC activity or expression of FCMF, particularly disorders of cell
CC survival, differentiation and proliferation, e.g. cancer and

CC preclampsia.
XX
SQ Sequence 71 AA:

Query Match 24.4%; Score 95; DB 20; Length 71;
Best Local Similarity 37.3%; Pred. No. 0.0051;
Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

OY 1 CEEFVCPQGGSDCDINDGPPVCRDRCEKEPS--FTCASDGLTYNNRCYMDAEACLRG 57
Db 3 cdgvec-gp9kacrmjgiprcecapcsglparlqvcsgdgatcydecelraarc-rg 59

RESULT 9

AAI14571
ID AAY14571 standard; Protein: 263 AA.

AAV14571;

14-SEP-1999 (first entry)

Human apoptosis inhibitory factor PCRF35 protein.

KW Neotrophic; neuroprotective; apoptosis inhibitory factor; PCRF35; rat;
KW pheochromocytoma cell; PC12 cell; human; bcl12; nerve cell death;
KW Parkinson's disease; Alzheimer's disease; human.

OS Homo sapiens.

PN WO9931237-A1.

24-JUN-1999.

11-DEC-1998; 98WO-JP05609.

14-MAY-1998; 98JP-0131634.

12-DEC-1997; 97JP-0343112.

(TAIS) TAISHO PHARM CO LTD.

PI Ohsawa Y, Uchiyama Y;

WPI: 1999-395180/33.

N-PSDB; AAX83200.

CC New protein obtained by culturing pheochromocytoma transformed by a
CC bcl2 gene as a protooncogene, useful in remedies for
CC apoptosis-caused diseases. e.g. Alzheimer disease

Example 4; Page 45-46; 50pp; Japanese.

CC This sequence represents the human apoptosis inhibitory factor PCRF35.

CC The coding sequence was isolated from a human foetal brain cDNA library

CC using the sequence of the corresponding rat gene (AAX79369) as a probe.

CC The protein can be used for the treatment of apoptosis-caused diseases,

CC including those due to nerve cell death, e.g. Parkinson's disease and

CC Alzheimer's disease.
XX
SQ Sequence 263 AA:

Query Match 24.4%; Score 95; DB 20; Length 263;
Best Local Similarity 37.3%; Pred. No. 0.02;
Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

OY 1 CEEFVCPQGGSDCDINDGPPVCRDRCEKEPS--FTCASDGLTYNNRCYMDAEACLRG 57
Db 99 cdgvec-gp9kacrmjgiprcecapcsglparlqvcsgdgatcydecelraarc-rg 155

RESULT 10

AAV06201

QY 1 CEEFVCPQGSDCDINDGQPVCRCRDRCEKPS--FTCASDGLTYYNRCYMDAEACLRG 57
 | : | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dd 99 cdgvec-gpykacrmlygrprcecapdcsglparlqvcsdgatylrdecelrarcc-rg 155

Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

QY 1 CEGFVCPQGSDCDIMDGGPVCRCRDRCEKPS--FTCASDGLTYVNRCYMDAEACLRG 57
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 99 cdgvec-gpgkacrmlygrprcecapdcsglparlvqcgsgdaltyrdecclnarc-rg 155

```

E17 315 310 /label= osteonec1n_domain
E18 315 310

```

FT	Modified-site	215..219
----	---------------	----------

SD Sequence 263 AA;

Query Match	24.48;	Score 95;	DB 22;	Length 263;
Best Local Similarity	37.39;	Prod No 0.03;		

Best Local Similarity	37.3%	Best Local Similarity	28.2%	Zeng et al., 2003
Matches	22; Conservative	6; Mismatches	27; Indels	4; Gaps

OY 1 CEGFVCPOOGSCDDIWDGPGVCRCHDRKEEPS--FTCASDGLTYNNRCYMDAEACTNG 57
 | : | | : : | : | : | : | : | : | :
Db 99 cdgvec-gpgkacimlgsfprcecapdcsqjparlqvcsdgatyrdecelnarcc-rg 155

Search completed: February 26, 2002, 01:26:41
Job time: 315 sec

[illegible]

Search completed: February 26, 2002, 01:26:41
Job time: 315 sec

Search completed: February 26, 2002, 01:26:41
Job time: 315 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:16 ; Search time 144.8 Seconds
(without alignments)
120.210 Million cell updates/sec

Title: US-09-819-136-2_COPY_39_157

Perfect score: 692

Sequence: 1 PNLWDASTGECRSRDOD.....RCYMDEACGLGLHLYPC 119

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Database : Listing first 45 summaries

SPREMBL.17:*
1: SP_Archea:*
2: SP_Bacteria:*
3: SP_Fungi:*
4: SP_Human:*
5: SP_Invertebrate:*
6: SP_Mammal:*
7: SP_mhc:*
8: SP_Organelle:*
9: SP_Phage:*
10: SP_Plant:*
11: SP_Rodent:*
12: SP_Virus:*
13: SP_Vertebrate:*
14: SP_Unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	19.2	1296	5 Q22452	022452 caenorhabd
2	125.5	18.1	319	13 Q91376	091376 xenopus lae
3	121.5	17.6	2026	4 Q00468	000468 homo sapien
4	120.5	17.4	343	13 Q90844	090844 gallus gall
5	120.5	17.4	343	13 Q9PS97	09PS97 gallus gall
6	117.5	17.0	195	5 Q9NDT4	09NDT4 balanus amp
7	117.5	17.0	299	13 Q9PCB0	09PCB0 xenopus lae
8	116.5	16.8	344	4 Q9BTH0	09BTH0 homo sapien
9	116	16.8	315	13 Q9W600	09W600 gallus fasc
10	115.5	16.7	308	6 Q9GKY0	09GKY0 macaca fasc
11	115	16.6	654	5 Q9V5J9	09V5J9 drosophila
12	111.5	16.1	256	11 Q99PW7	099PW7 rattus norv
13	111.5	16.1	306	11 Q99J19	099J19 mus musculu
14	111.5	16.1	3367	5 Q9XCC9	09XCC9 drosophila
15	110.5	16.0	282	13 Q9S988	09S988 salveinus
16	109.5	15.8	3680	5 Q9VR08	09VR08 drosophila
17	108.5	15.7	2150	5 Q44131	044131 caenorhabd
18	107.5	15.5	256	11 Q9BQC7	09BQC7 mus musculu
19	105.5	15.2	121	13 Q91450	091450 salveinus

20	105.5	15.2	322	13 Q9YHV4	09YHV4 brachydanto
21	104.5	15.1	1511	5 Q9VB21	09VB21 drosophila
22	103.5	15.0	4441	5 Q9W5X1	09W5X1 drosophila
23	103.5	15.0	2524	5 Q9GPAS	09GPAS brachyosteo
24	103	14.9	3319	5 Q9VJ75	09VJ75 drosophila
25	101.5	14.7	263	4 Q9S633	09S633 homo sapien
26	100	14.5	1530	11 Q9WUG5	09WUG5 rattus norv
27	100	14.5	1531	11 Q88279	088279 rattus norv
28	99.5	14.4	664	4 Q9U117	09U117 homo sapien
29	99.5	14.4	657	11 Q9R158	09R158 mus musculu
30	99.5	14.4	1534	4 Q75093	075093 homo sapien
31	99	14.3	1531	11 Q9WVB5	09WVB5 mus musculu
32	98.5	14.2	903	5 Q44397	044397 trichuris t
33	98.5	14.2	1162	5 Q9S011	09S011 caenorhabd
34	98.5	14.2	1523	11 Q88280	088280 rattus norv
35	98.5	14.2	2703	5 Q9W4T8	09W4T8 drosophila
36	98.5	14.2	2704	5 Q97458	097458 drosophila
37	97.5	14.1	619	4 Q9H1F8	09H1F8 homo sapien
38	97.5	14.1	1523	4 Q75094	075094 homo sapien
39	96.5	13.9	715	5 Q94494	094494 dictyostell
40	96.5	13.9	2717	5 Q94710	094710 paramecium
41	95.5	13.8	2471	11 Q9QW30	09QW30 rattus sp.
42	95	13.7	346	4 Q9NS55	09NS55 homo sapien
43	95	13.7	368	4 Q9P2Y9	09P2Y9 homo sapien
44	95	13.7	374	4 Q9UK5	09UK5 homo sapien
45	94.5	13.7	850	4 Q9ULF7	09ULF7 homo sapien

ALIGNMENTS

RESULT 1
ID Q22452 PRELIMINARY; PRT; 1296 AA.
AC Q22452;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SMILAR TO AGRIN AND FOLLISTATIN.
GN T13C2.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditiolidae;
OC Rhabditiidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID:6239;
RN [1]
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Shalton N., Smith A., Sonhammer E., Straden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Du Z.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U40030; AAA81133.1; -.
DR HSSP: P37109; 1PCE.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; Kazal.

SEQUENCE 2026 AA; 212881 MW

Query Match	17.6%	Score 121.5	DB 4	Length 2026
Best Local Similarity	24.4%	Pred. No. 7.4e-05		
Matches	38	Conservative	12	Mismatches 47; Indels 59; Gaps 8
OY	14	EC-SRDQCCAAAEKCCINVCGLHSCVAAAFPGSPAFTTAAASCEGF	---	VCPOGS--DC 67
Db	369	OCQGRDQ---	CEPRFRFNAV---	CLSR--GRPGSCDRVTCDCGAYRYCAGDGRFTYS 419
OY	68	DIW-----		DGQVRCRDRCEKEP 86
Db	420	DCWRQOAECRQORAIAPSKHOGCPDQAPSECLGVCQAFGATCAVKNGQAAACECLQACSSLY		479
OY	87	SETGASDGLTYYNRCYMDAEACRLGLHLHIY	---	PC 119
Db	480	DPVCGSDGYTIGSACELENTACTLGREIOWARKGPC		515
RESULT	4			
ID	090844	PRELIMINARY:	PRT:	343 AA.
AC	090844;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	FOLLISTATIN.			
OS	Gallus gallus (chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
CC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96189336; PubMed=8625798;			
RA	Graham A., Lumsden A.;			
RT	"Interactions between rhombomeres modulate Krox-20 and follistatin			
RT	expression in the chick embryo hindbrain.";			
RL	Development 122:473-480(1996).			
DR	EMBL: X87609; CAA60915.1; .			
DR	HSP: P00998; ITGS.			
DR	InterPro: IPR003645; FOLN.			
DR	InterPro: IPR002350; kazal.			
DR	PIfam: PF00050; kazal; 3.			
DR	SMART: SM00274; FOLN; 3.			
DR	SMART: SM00280; KAZAL; 3.			
SO	SEQUENCE	343 AA;	38192 MW;	15A78762560ELAE CRC64.
Query Match	17.4%	Score 120.5	DB 13	Length 343;
Best Local Similarity	25.6%	Pred. No. 2.1e-05		
Matches	41	Conservative	12	Mismatches 60; Indels 47; Gaps 8
OY	1	PNLWVDAOSTEBRESRQDCAAEKCCINVCGLHSCVAA-RFPGSPAAP	-----	49
Db	153	PELEVQYGKCKKTC-RDVLCPGSSCTCYVDQTNNAVYCYCNRIICEPPTSPEQYLCGNDGI		211
OY	50	-----	TTASCEGFVCPQGSDDDIWD	---GQVCR-C 78
Db	212	TYASACHIRKATCLIGRISIGLAIECKCIKASCEIDIQ-SAGKKC-LMDFKVGGRGALC		269
OY	79	RDRC--EKEPSFTCASDGLTYYNRCYMDAEACRLGLHLHI		116
Db	270	DELCPESKSDAVACASDNTYTPSECCAMKEAASMGVLLLEV		309
RESULT	5			
ID	09PS97	PRELIMINARY:	PRT:	343 AA.
AC	09PS97;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	FOLLISTATIN-NEURAL INDUCER.			

Query Match	17.4%	Score 120.5	DB 13	Length 343
Best Local Similarity	25.6%	Pred. No. 2.1e-05		
Matches 41	Conservative 12	Mismatches 60	Indels 47	Gaps 8
<p>OS Gallus gallus (chicken).</p> <p>OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;</p> <p>OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;</p> <p>OC Gallus.</p> <p>OX NCBI_TaxID=9031;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>FX MEDLINE=96029008; PubMed=7554497;</p> <p>RA Connolly D.J., Patel K., Seletro E.A., Wilkinson D.G., Cooke J.;</p> <p>RT "Cloning, sequencing, and expressional analysis of the chick homologue</p> <p>of follistatin.";</p> <p>RL Dev. Genet. 17:65-77(1995).</p> <p>DR HSSP: P00998; IrgS.</p> <p>DR InterPro: IPR003645; FOLN.</p> <p>DR InterPro: IPR002350; kazal.</p> <p>DR Pfam: PF00050; kazal; 3.</p> <p>DR SMART: SM00274; FOLN; 3.</p> <p>DR SMART: SM00280; KAZAL; 3.</p> <p>SO SEQUENCE 343 AA; 38182 MW; A92B5699DDAE5ECT CRC64;</p>				
<p>Query Match 17.4% Score 120.5; DB 13; Length 343;</p> <p>Best Local Similarity 25.6%; Pred. No. 2.1e-05;</p> <p>Matches 41; Conservative 12; Mismatches 60; Indels 47; Gaps 8;</p>				
<p>OY 1 PNLVDAOSTERECRSRDQCAAEEKCCINWGLHSCVAA-RFPSPAP-----49</p> <p>DB 153 PELEVOYGKCKKTC-RVLCGSSSTCVADQLNNAYCVTCNRCICEPTSPGYLGNDGI 211</p> <p>OY 50 -----TTASCEGFVPOQSDDDIND--GQYVCR-C 78</p> <p>DB 212 TYASACHLRKATCILLGRSIGLAYPEGCKIKANSCEDIQC-SAGKC-LMDFKVRGRCALC 269</p> <p>OY 79 RDRC--EKEPSFTCASDGLTYNRCYMAEACLRGLHI 116</p> <p>DB 270 DELPEKSDAEAVCASDNTTTPSECCAMKAEACSMGLLEV 309</p>				
RESULT 6				
ID 09NDT4	PRELIMINARY;	PRT:	195 AA.	
AC 09NDT4;				
DT 01-OCT-2000 (TREMBLrel. 15, Created)				
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE BCS-4.				
GN BCS-4.				
GN Balanus amphitrite (Barnacle).				
OS Eukaryota; Metazoa; Arthropoda; Crustacea; Maxillopoda; Cirripedia;				
OC Thoracica; Sessilia; Balanomorpha; Balanoidae; Balanidae; Balanus.				
OX NCBI_TaxID=32267;				
OX [1]				
RN SEQUENCE FROM N.A.				
RP MEDLINE=20314484; PubMed=10854786;				
RA Okazaki Y., Shitani Y.;				
RT "Structures of six cDNAs expressed specifically at cypris larvae of				
baroncles, Balanus amphitrite.";				
RL Gene 250:127-135(2000).				
DR EMBL: AB021905; BAA9546.1; -				
DR InterPro: IPR003645; FOLN.				
DR SMART: SM00274; FOLN; 2.				
SO SEQUENCE 195 AA; 20545 MW; 72204A9304191EDI CRC64;				
<p>Query Match 17.0% Score 117.5; DB 5; Length 195;</p> <p>Best Local Similarity 32.5%; Pred. No. 2.7e-05;</p> <p>Matches 38; Conservative 9; Mismatches 47; Indels 23; Gaps 7;</p>				
OY 6 DAOSICERECSDDOC---AAAEKCCINWGLHSCVAA---ARFGSPAAPTAAECGF 58				
DB 60 EAPSKTKKCSBDCCEPTPAGPVYQXTCADIKCAAGICESELPDGPCCVDEVSCSGF 119				
OY 59 VCPQGSQCDIMDGPVC---RCRDR-CEKEPSFTCASDGLTYNRCYMAEACLRG 111				

Db 120 FC-ORGINCYLRDGSPTCLPNTCEVRECCK-----GLTCLDT--PDGALCRG 164

RESULT 7

09DGB0 PRELIMINARY; PRT: 299 AA.
AC 09DGB0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN.
GN XERP.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Okabayashi K., Shoji H., Asashima M., Sugino H.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99121187; PubMed=9920730;
RA Okabayashi K., Shoji H., Onuma Y., Nakamura T., Nose K., Sugino H.,
RT "CDN Cloning and Distribution of the Xenopus Follistatin-Related
RT Protein."
RT Biochem. Biophys. Res. Commun. 254:42-48(1999).
CC -1- SIMILARITY: NO EF-HAND FAMILY.
DR EMBL: AB049354; BAB13800.1; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003645; FOLN.
DR Pfam: PF00036; efhand.2.
DR SMART: SM00050; kazal.1.
DR SMART: SM00274; FOLN; 1.
DR SMART: SM00280; KAZAL; 1.
KW Calcium-binding.
SQ SEQUENCE 299 AA; 33701 MW; 43ADEF066F5E62AF CRC64;

Query Match 17.0%; Score 117.5; DB 13; Length 299;
Best Local Similarity 25.0%; Pred. No. 3.9e-05;
Matches 26; Conservative 14; Mismatches 33; Indels 31; Gaps 3;

OY 15 CSRDPCAAREKCCINVCGLHSCVAARPPGSPAFTTAASCEGFVCPQGSDDCDIMD-G-72
DB 15 CALEPRKSKSKVCANVFCG-----AGRECAVTEKG 45
OY 73 QPVCRGRDCEKEPSFTCASDGLTYNRCYMDAECRLGLHLHI 116
DB 46 DPTCLCIERCKSHKRPVCGSGNGKTYLNLHCELHRDACLGTSGKIYV 89

RESULT 8

09BTH0 PRELIMINARY; PRT: 344 AA.
AC 09BTH0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FOLLISTATIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC ADENOCARCINOMA;
RA Strauberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC004107; AAH04107.1; -
SQ SEQUENCE 344 AA; 38007 MW; D9BB45055D84AC90 CRC64;

Query Match 16.8%; Score 116.5; DB 4; Length 344;
Best Local Similarity 25.9%; Pred. No. 5.7e-05;
Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8;

OY 1 PNLVMDAOSTCEKEPSFTCASDGLTYNRCYMDAECRLGLHLHI 49
DB 154 PELEVOYGRCKRTCC-RDYFCPSGSGTCVVDQTNNAVCYCNCRICPPSPSEGYLCNDGV 212
OY 50 -----TTAASCEGVCPQGSDDCDIMD---GQPVCR-- 77
DB 213 TYSSACHLRKATLLRSGISGLAVEGKICIRAKSCEDIQC-TGGRKC-LMPFKYGRGRCSLC 270
OY 78 ---CRDRCEKEPSFTCASDGLTYNRCYMDAECRLGLHLHI 116
DB 271 DELCPDSKSDP--VCASDNTATASCAKMEACSSGVLLV 310

RESULT 9

09W600 PRELIMINARY; PRT: 315 AA.
AC 09W600;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FLIK PROTEIN.
GN FLIK.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Patel K.;
RT "Cloning and early dorsal-axial expression of FLIK, a chick
RT follistatin-related gene: evidence for involvement in
RT dorsalisation/neural induction."
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96427388; PubMed=8812133;
RA Patel K., Connolly D., Amthor H., Nose K.;
RT "Cloning and early dorsal axial expression of FLIK, a chick
RT follistatin-related gene: evidence for involvement in
RT dorsalisation/neural induction."
RT Dev. Biol. 178:327-342(1996).
DR EMBL: AJ238977; CAB42968.1; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003645; kazal.
DR InterPro: IPR003645; FOLN.
DR Pfam: PF00050; kazal.1.
DR SMART: SM00274; FOLN; 1.
DR SMART: SM00280; KAZAL; 1.
SQ SEQUENCE 315 AA; 35815 MW; C01889E005658A67 CRC64;

Query Match 16.8%; Score 116; DB 13; Length 315;
Best Local Similarity 27.6%; Pred. No. 6e-05;
Matches 24; Conservative 18; Mismatches 41; Indels 4; Gaps 3;

OY 31 VCGHSCVAARFPSPAPFTTAASCEGFVCPQGSDDCDIMD-GQPVCRDRDCEKEPSFT 89
DB 9 LCALLAARLRAEPEPSKSKI--CANVFC-GRGAECAVTEKEGPTCLTEQCKPHRPV 65
OY 90 CASDGLTYNRCYMDAECRLGLHLHI 116
DB 66 CGSNGKTYLNLHCELHRDACLGTSGKIYV 92

```

RESULT 10
O9GKY0 PRELIMINARY: PRT: 308 AA.
AC O9GKY0:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OCC1.
GN OCC1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecidae; Macaca.
OX NCBI_TaxId=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Tochtiani S.;
RT "Neuronal heterogeneity in macaque neocortex revealed by occ1 mRNA
activity-dependent expression.";
RT Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.
CC -1. SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AB039661; BAB20770.1;
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; KAZAL.
DR InterPro: IPR001007; WFC.
DR Pfam: PF00036; ehand.2.
DR Pfam: PF0050; kazal.1.
DR SMART: SM00274; FOLN; 1.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00214; WVC; 1.
DR SMART: SM00011; WVC_def; 1.
KW Calcium-binding.
SQ SEQUENCE 308 AA; 34999 MW; 4B2836D9CDF535D9 CRC64;

```

```

Query Match 16.7%; Score 115.5; DB 6; Length 308;
Best Local Similarity 23.5%; Pred. No. 6.6e-05;
Matches 27; Conservative 16; Mismatches 31; Indels 41; Gaps 4;

```

```

OY 4 WDAOSTCERESRDQCAAEKCCIN-CGLHSCVARFPSPAPPTAASCEGFVCPQ 62
DB 17 WVR-----EELNSKSKICANVFC----- 37
OY 63 QSSCDIWD-GQPVRCRCRDRCEKPEFTCASDGLTYNNRCYMDAECALGHLHI 116
DB 38 AGRECAVTEKGEPCTLCIECKPKHRRPVCGSNGKTYLHNCHELRDACLGTGSKIOW 92

```

```

RESULT 12
O99PW7 PRELIMINARY: PRT: 256 AA.
AC O99PW7:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN FLRG PRECURSOR.
GN FLRG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC Uchiyama Y.; Ohsawa Y.; Kametaka S.;
RA "rat follistatin-related protein FLRG (rat PCTF35, PC12 cell derived
trophic factor 35).";

```

```

Query Match 16.6%; Score 115; DB 5; Length 654;
Best Local Similarity 27.7%; Pred. No. 0.00014;
Matches 33; Conservative 9; Mismatches 51; Indels 26; Gaps 5;

```

```

OY 11 CERESRDQCAAEK-----CINYGLHSCVARFPSPAPPTAASCEGFVCPQ 64
DB 156 CPRSCPSTIVGAEVPVCGSLYANICELRKTKCSR-SGVSLIKVDRCCE-----RSKG 210
OY 65 SDCDINDGQPVRCRCRDRCEKPEFTCASDGLTYNNRCYMDAECALGHLH----HYPC 119
DB 211 SD-----CKHRCSTEKDPVCGTGDRTYLNRCMLRVOSCRVGLAAVKLSHWPC 258

```

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB021295; BAB32664.1; -
 KW Signal.
 FT SIGNAL
 SO SEQUENCE 256 AA; 27109 MW; C846416C0A1D56F CRC64;

Query Match 16.1%; Score 111.5; DB 11; Length 256;
 Best Local Similarity 26.8%; Pred. No. 0.00015;
 Matches 34; Conservative 15; Mismatches 39; Indels 39; Gaps 7;

QY 8 QSC-----EKEGRDDCAAEKCCINVC-----GLHSCVAARPPGSP 46
 DB 43 EATCSLVLTQVYSRECCASGN---INTAMSNFTHPGNKISILIGFLVHCIPCCK----- 94
 QY 47 AAPTASGEFVCPQGGSDCDIMDGPVRCRDRCRCEKPPS--FTCASGLTYNNCYMD 104
 DB 95 -----DSCDGVCC-CPGACRMLGSRPHCEVNCCEGVAGFQVCGSDGATYRDECELR 147
 QY 105 AEAQCLNG 111
 DB 148 TARC-RG 153

RESULT 13
 Q99J19 PRELIMINARY; PRT; 306 AA.
 AC Q99J19;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE POLIISTATIN-LIKE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC006185; AA06185.1; -
 SO SEQUENCE 306 AA; 34554 MW; 0AEB11C621B8FDC CRC64;

Query Match 16.1%; Score 111.5; DB 11; Length 306;
 Best Local Similarity 23.8%; Pred. No. 0.00018;
 Matches 24; Conservative 16; Mismatches 30; Indels 31; Gaps 3;

QY 18 DQDCAAEKCCINVCGLHSCVAARFPSPAPPTAASCEGFVCPQGGSDCDIMD-CQPV 75
 DB 19 EEPFRSKKICANVPG-----AGRECAVTEKEGPT 49
 QY 76 CRCHDRCKEPEFTCASDGLTYNNCYMDAECRLGLHLHI 116
 DB 50 CLCIEOCKPHRRPVCGSNKTYLNHCILHRDACLGLSKIQV 90

RESULT 14
 Q9XZC9 PRELIMINARY; PRT; 3367 AA.
 AC Q9XZC9;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LAMININ ALPHA1.2 (SYMBOL-WB).
 GN WB OR WING BLISTER OR BG:DS03792.1 OR CG15288.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;
 RX MEDLINE=99207061; PubMed=10189378;
 RA Martin D., Zusman S., Li X., Williams E.L., Khare N., Parocha S.,
 RA Chiquet-Ehrismann R., Baumgartner S.;
 RT "wing blister, a new Drosophila laminin alpha chain required for cell
 RT adhesion and migration during embryonic and imaginal development.";
 RL J. Cell Biol. 145:191-201(1999).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Y, AND CN BW SP;
 RX MEDLINE=99403001, PubMed=10471707;
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazey R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshirefi A.,
 RA Palazolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitehaw K.,
 RA Celniker S., Rubin G.M.;
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.";
 RL Genetics 153:179-219(1999).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=Y, AND CN BW SP;
 RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazey R.G.,
 RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomolan M.A., Mazda P., Moshirefi A.R., Moshirefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
 RA Setchell H., Smit E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF135118; AD031714.1; -
 DR EMBL; AE003409; AAP44861.1; -
 DR HSP; P00740.1EDM; -
 DR FlyBase; FBgn0004002; wb.

DR InterPro; IPR001589; Actinin_act_bind.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR001866; LamNT.
 DR Pfam; PF00052; Laminin_B; 1.
 DR Pfam; PF00053; Laminin_EGF; 15.
 DR Pfam; PF00054; Laminin_G; 4.
 DR Pfam; PF00055; Laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF-LAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR ProDom; PD003031; Laminin_B; 2.
 DR SMART; SM00180; EGF_Lam; 16.
 DR SMART; SM00281; Lamb; 2.
 DR SMART; SM00282; Lamb; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_14.
 DR PROSITE; PS01166; EGF_2; 4.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
 DR EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.
 SO SEQUENCE 3367 AA; 374097 MW; EB12654B1B01511 CRC64;

Query Match 16.1%; Score 111.5; DB 5; Length 3367;
 Best Local Similarity 24.7%; Pred. No. 0.0014;
 Matches 37; Conservative 19; Mismatches 41; Indels 53; Gaps 11;

QY 9 STERECSRDQDC---AAAEKCCINVCGLHSCVAARFPSPAPPTAASCEGFVCPQGG 65
 DB 1677 SDDDRREGVQCNCRGNTGSDHC-----HQE-AEGFYDPPNP---HCCQACPCPETNR 1725
 QY 66 D-----CDIMDGPVRCR-----DRCE-----KEPSFTCA-----SDGL----- 95
 DB 1726 NFARGCNVMDGEVSCVCKRGYTGRLCERCQAGYFGDPMPYPTNTQPCPCNCHPDGIDTBOC 1785
 QY 96 -TYNNRCY-----WDARECLRLGLHLHIY 117

Db 1786 DVETGRGCRGCVTGLKCDKCGAERH-HLV 1814

RESULT 15

Q98988 PRELIMINARY; PRT: 262 AA.
AC Q98988:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OVIOLATORY PROTEIN-2 PRECURSOR.
OS Salvelinus fontinalis (Brook trout) (Brook char).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Plectacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_Taxid=8038;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA Garczynski M.A.; Goetz F.W.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U67854; AAB63598.1; -
DR HSSP; P19957; 1FLE.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00095; WAP; 4.
DR PRINTS; PRO0003; 4DISULPHORE.
DR PRODOM; PD01224; WAP; 2.
DR SMART; SM00217; WAP; 3.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 5.
KW Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 262 OVULATORY PROTEIN-2.
SQ SEQUENCE 262 AA; 28200 MM; D2EC50FFAFECF807 CRC64;

Query Match 16.0%; Score 110.5; DB 13; Length 262;

Best Local Similarity 25.4%; Pred. No. 0.0002; Mismatches 49; Indels 43; Gaps 8;

QY 10 TCERCESRDODCAAEKCCINVCGLHSCVAFPGSPAAPTAAAC-----EGFV 59
Db 93 TCAELCSDDSDCPNDKCKCHNGC-HVCIA---PYTVVLKAGCCPLLLKVP SHKGCY 148
QY 60 ----CPQGSDDIWDGQPC-----RCRDCEKEPSEFTCASDGLT 96
Db 149 RDEDCPKD-DKCVFHSDALCVPPDFTKPGVCPRRRMGVGTCAELCSNDS--CPNDEKC 205
QY 97 YNRCYMDAEACLRGLHLIHP 118
Db 206 CHNGCGHD--CIAPYVPIIP 224

Search completed: February 26, 2002, 01:49:18
Job time: 1267 sec

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:37 : Search time 45.78 Seconds

(without alignments)
95.306 Million cell updates/sec

Title: US-09-819-136-2_COPY_39_157

Perfect score: 692
Sequence: 1 PNLWDASTGECRSRQD.....RCYMDAENACRLGHLHIYVC 119

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.5	18.1	341	1 FSA_XENLA	P31515 xenopus lae
2	121.5	17.6	1328	1 AGRI_DISOM	O90404 discopys o
3	120.5	17.4	343	1 FSA_CHICK	O90844 gallus gall
4	120.5	17.4	344	1 FSA_HORSE	O62650 equus caball
5	116.5	16.8	338	1 FSA_HUMAN	P19883 homo sapien
6	115.5	16.7	308	1 FSLI_HUMAN	Q12841 homo sapien
7	115.5	16.7	308	1 FSLI_HUMAN	O99X00 macaca fasc
8	115.5	16.7	1955	1 AGRI_CHICK	P31696 gallus gall
9	113.5	16.4	337	1 FSA_SHEEP	P31514 ovis aries
10	113.5	16.4	344	1 FSA_MOUSE	P47931 mus musculu
11	113.5	16.4	344	1 FSA_PIG	P10669 sus scrofa
12	113.5	16.4	344	1 FSA_RAT	P21674 rattus norv
13	112	16.2	1959	1 AGRI_RAT	P25304 rattus norv
14	111.5	16.1	306	1 FSLI_MOUSE	O62356 mus musculu
15	111.5	16.0	344	1 FSA_BOVIN	P50291 bos taurus
16	110.5	15.3	1429	1 FSA_BRARE	P07207 drosophila
17	106	15.3	1429	1 FSA_BRARE	O9YH44 brachydanio
18	105.5	15.2	2703	1 NOTC_DROME	P07207 drosophila
19	98.5	14.2	619	1 MTNA_HUMAN	O95460 homo sapien
20	97.5	14.1	790	1 AD30_HUMAN	O9UKE2 homo sapien
21	96	13.9	134	1 AD15_HUMAN	O13444 homo sapien
22	96	13.6	134	1 EPTI_MOUSE	O96A01 mus musculu
23	94	13.6	712	1 FBLI_CAEL	O77469 caenorhabdi
24	91.5	13.2	726	1 AD20_HUMAN	O43506 homo sapien
25	91.5	13.2	1696	1 PCRS_BRACL	O9N155 brachylosto
26	91.5	13.2	4543	1 LRP1_CHICK	P98157 gallus gall
27	91.5	13.2	4543	1 LRP2_HUMAN	P98157 gallus gall
28	91.5	13.2	5376	1 ZAN_MOUSE	O88739 mus musculu
29	91.5	13.2	803	1 ITBI_CHICK	P07228 gallus gall
30	91	13.2	1107	1 YLR2_CAEL	P41950 caenorhabdi
31	91	13.2	1192	1 LMG2_MOUSE	O61092 mus musculu
32	90.5	13.1	4544	1 LRP1_HUMAN	O07954 homo sapien
33	90	13.0			

ALIGNMENTS

RESULT	1	STANDARD	PRT	341 AA.
FSA_XENLA				
ID	P31515	O91376		
AC	01-JUL-1993 (Rel. 26, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN) (XFS-319).			
GN	FST.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RP	SEQUENCE FROM N.A. (SHORT ISOFORM), AND CHARACTERIZATION.			
RC	TTSUE-Tail bud;			
RC	MEDLINE=94221645; PubMed=8168135;			
RA	Hemmati-Briavaniou A., Kelly O.G., Melton D.A.;			
RT	"Follistatin, an antagonist of activin, is expressed in the Spemann			
RT	organizer and displays direct neutralizing activity.";			
RL	Cell 77:283-295(1994).			
RN	[2]			
RP	PARTIAL SEQUENCE FROM N.A. (LONG ISOFORM).			
RC	TTSUE-Ovary;			
RC	MEDLINE=91128357; PubMed=1704219;			
RA	Teshiro K., Yamada R., Asano M., Hashimoto M., Muramatsu M.,			
RA	Shiohara K.;			
RT	"Expression of mRNA for activin-binding protein (follistatin) during			
RT	early embryonic development of Xenopus laevis.";			
RL	Biochem. Biophys. Res. Commun. 174:1022-1027(1991).			
RP	SEQUENCE OF N-TERMINUS, AND CHARACTERIZATION.			
RX	PubMed=8365557;			
RA	Fukui A., Nakamura T., Sugino K., Takio K., Uchiyama H., Aeshima M.,			
RA	Sugino H.;			
RT	"Isolation and characterization of Xenopus follistatin and activins.";			
RL	Dev. Biol. 159:131-139(1993).			
CC	-1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN			
CC	ANTAGONIST WHICH PLAYS A ROLE IN NEURAL INDUCTION. THE SHORT			
CC	ISOFORM IS A MORE POTENT INHIBITOR OF ACTIVIN THAN THE LONG			
CC	ISOFORM. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION OF			
CC	PITUITARY FOLLICLE STIMULATING HORMONE (FSH).			
CC	-1- SUBUNIT: MONOMER (POTENTIAL).			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A			
CC	SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	-1- TISSUE SPECIFICITY: SPERMATOGONIAL AND NOTOCHORD.			
CC	-1- DEVELOPMENTAL STAGE: THE SHORT ISOFORM IS PRESENT INTERNALLY WHILE			
CC	THE LONG ISOFORM IS EXPRESSED AT GASTRULA STAGES. DETECTED IN A			
CC	FEW CELLS OF THE SPERMATOGONIAL AT THE ONSET OF GASTRULATION.			
CC	DURING GASTRULATION EXPRESSION CONTINUES IN THE PRECHORDAL PLATE			
CC	AND THE ANTERIOR PORTION OF THE NOTOCHORD ANlage. BEGINNING AT			
CC	EARLY NEURULA STAGES, EXPRESSION IS INITIATED AT NEW STAGES IN THE			
CC	HEAD MESODERM; HYPOCHORD; PRONEPHROS; EYES; FORE-, MID-, AND			
CC	HINDBRAIN; AND THE MIDBRAIN-HINDBRAIN JUNCTION.			

34	89.5	12.9	788	1 ITB3_HUMAN	P05106 homo sapien
35	89	12.9	1700	1 BAR3_CHITE	O03376 chironomus
36	89	12.9	1964	1 NTC4_MOUSE	P31695 mus musculu
37	89	12.9	4289	1 TBNX_HUMAN	P22105 homo sapien
38	88.5	12.8	787	1 ITB3_MOUSE	O54890 mus musculu
39	88.5	12.8	820	1 AD29_HUMAN	O9UKF5 homo sapien
40	88.5	12.8	1607	1 LMG1_MOUSE	P02468 mus musculu
41	88	12.7	798	1 ITB1_FELCA	P53713 felis silve
42	88	12.7	798	1 ITB1_MOUSE	P09055 mus musculu
43	88	12.7	2318	1 NTC3_MOUSE	O61962 mus musculu
44	87.5	12.6	191	1 WAP_MACBU	O9N018 macropus eu
45	87.5	12.6	676	1 KALM_CHICK	P33005 gallus gall

```

CC -1- INDUCTION: BY ACTIVIN.
CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: S69801; AAB30638.1; -
DR PIR: JN0119; JN0119.
DR HSSP: P37109; 1PCE.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00050; kazal; 3.
DR SMART: SM00274; FOLN; 3.
DR SMART: SM00280; KAZAL; 3.
DR PROSITE: PS00282; KAZAL; FALSE_NEG.
KW Glycoprotein; Repeat; Signal; Alternative splicing.
FT SIGNAL 1 29
FT CHAIN 30 341
FT DOMAIN 94 117 FOLLISTATIN 1.
FT DOMAIN 118 164 KAZAL-LIKE 1.
FT DOMAIN 167 190 FOLLISTATIN 2.
FT DOMAIN 192 239 KAZAL-LIKE 2.
FT DOMAIN 244 268 FOLLISTATIN 3.
FT DOMAIN 270 316 KAZAL-LIKE 3.
FT DOMAIN 324 331 POLY-GLO.
FT DISULFID 118 150 BY SIMILARITY.
FT DISULFID 122 143 BY SIMILARITY.
FT DISULFID 132 164 BY SIMILARITY.
FT DISULFID 192 225 BY SIMILARITY.
FT DISULFID 196 218 BY SIMILARITY.
FT DISULFID 207 239 BY SIMILARITY.
FT DISULFID 270 302 BY SIMILARITY.
FT DISULFID 274 295 BY SIMILARITY.
FT DISULFID 284 316 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VANSPLIC 318 341 SIVETEEEEEEDYSEFVLISSW -> CK (IN SHORT ISOFORM).
FT CONFLICT 268 268 A -> G (IN REF. 2).
FT CONFLICT 270 270 C -> S (IN REF. 2).
SQ SEQUENCE 341 AA; 37543 MW; 5D26E7BE70890171 CRC64;

```

```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AGRIN (FRAGMENT).
GN AGRN
OS Discopage ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Hypnosqualia; Pistiorajea; Batoidae;
OC Torpediniformes; Narcinoidae; Narcinidae; Discopage.
OX NCBI_Taxid-7785;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith M.A., Megill-Sole C., Rupp F., Yao Y.-M.M., Schilling J.W.,
RA Snow P., McMahon U.J.;
RT *Isolation and characterization of a cDNA that encodes an agrin
RT homolog in the marine ray.
RL Mol. Cell. Neurosci. 3:406-417(1992).
CC -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
CC JUNCTION (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS AT LEAST 2 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L01423; AAA49224.1; -
DR HSSP: P00740; 11XA.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF 2.
DR InterPro: IPR003884; FACI_MAC.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00008; EGF; 4.
DR Pfam: PF00050; kazal; 2.
DR Pfam: PF00053; laminin_EGF; 2.
DR Pfam: PF00054; laminin_G; 3.
DR Pfam: PF01390; SEA; 1.
DR SMART: SM00180; EGF_Lam; 2.
DR SMART: SM00001; EGF_Like; 3.
DR SMART: SM00057; FIMAC; 1.
DR SMART: SM00274; FOLN; 1.
DR SMART: SM00280; KAZAL; 2.
DR SMART: SM00282; Lam; 3.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; 5.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE: PS00024; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT NON_TER 1 1
FT DOMAIN 79 132 LAMININ EGF-LIKE 1.
FT DOMAIN 133 179 LAMININ EGF-LIKE 2.
FT DOMAIN 411 533 SEA.
FT DOMAIN 608 644 EGF-LIKE 1.
FT DOMAIN 865 902 EGF-LIKE 2.
FT DOMAIN 1097 1135 EGF-LIKE 3.
FT DISULFID 79 91 BY SIMILARITY.
FT DISULFID 81 98 BY SIMILARITY.
FT DISULFID 100 109 BY SIMILARITY.
FT DISULFID 112 130 BY SIMILARITY.

```



```
FT DISULFID 133 145 BY SIMILARITY.
FT DISULFID 135 152 BY SIMILARITY.
FT DISULFID 154 163 BY SIMILARITY.
FT DISULFID 166 177 BY SIMILARITY.
FT DISULFID 612 623 BY SIMILARITY.
FT DISULFID 617 632 BY SIMILARITY.
FT DISULFID 634 643 BY SIMILARITY.
FT DISULFID 869 880 BY SIMILARITY.
FT DISULFID 874 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 1101 1114 BY SIMILARITY.
FT DISULFID 1108 1123 BY SIMILARITY.
FT DISULFID 1125 1134 BY SIMILARITY.
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1328 AA; 144018 MW; 79D81C1AF2A71C18 CRC64;

Query Match 17.6%; Score 121.5; DB 1; Length 1328;
Best Local Similarity 27.6%; Pred. No. 0.00054;
Matches 40; Conservative 15; Mismatches 61; Indels 29; Gaps 9;

OY 1 PNTW-----VDAGSTEREC-----RDQCAAEKCCINVCGLHSCVAARF----- 43
DB 114 PGFNFGIYDEKSGC-TPCNCYPLGAVRD-DEQMSGLCSAGISGMKCNCPNGSK 171
OY 44 -----GSPAAPTASCEGFCVPOGSDCDIWDGQVPCRCRDR-CEKEPSF-TCASDGLTY 97
DB 172 LGPSGCGQDPSPVSRKTSGLHC-OYGATCVCOSIGRAYCECPSPICPKKNGFVCGSDGTY 230
OY 98 YNRCTYMDAEACLRGLHIV---PC 119
DB 231 ANECOLKTIACRQGSVINILHQCPC 255

RESULT 3
FSA_CHICK STANDARD; PRT; 343 AA.
ID FSA_CHICK Q90844; Q9P597; Q90680;
AC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
GN FST.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96189336; PubMed=8625798;
RA Graham A., Lumsden A.;
RT "Interactions between rhombomeres modulate Krox-20 and follistatin
expression in the chick embryo hindbrain.";
RL Development 122:473-480(1996).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96029008; PubMed=7554497;
RA Connolly D.J., Patel K., Seletro E.A., Wilkinson D.G., Cooke J.;
RT "Cloning, sequencing, and expression analysis of the chick homologue
of follistatin.";
RL Dev. Genet. 17:65-77(1995).
RN 13
RP SEQUENCE OF 177-250 FROM N.A.
RC TISSUE=Embryonic ovary;
RX MEDLINE=96049133; PubMed=7576634;
RA Darland D.C., Link B.A., Nishi R.;
```

```
RT "Activin A and follistatin expression in developing targets of ciliary
ganglion neurons suggests a role in regulating neurotransmitter
phenotype.";
RL Neuron 15:857-866(1995).
CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
ANTAGONIST. INHIBITS ACTIVIN A SIGNALING IN THE IRIS AND REGULATES
SOMATOSTATIN PHENOTYPE IN CILIARY GANGLION NEURONS. SPECIFIC
INHIBITOR OF THE BIOSYNTHESIS AND SECRETION OF PITUITARY FOLLICLE
STIMULATING HORMONE (FSH).
CC -1- SUBUNIT: MONOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: CILIARY GANGLION NEURONS. LEVELS ARE HIGHER IN
THE IRIS THAN THE CHOROID.
CC -1- DEVELOPMENTAL STAGE: LEVELS INCREASE IN THE IRIS FROM EMBRYONIC
DAY 9 (E9) TO E16 IN CONTRAST TO THE CHOROID WHERE IT REMAINS LOW
RELATIVE TO IRIS. DURING EARLY HINDBRAIN DEVELOPMENT STRONGLY
EXPRESSED IN RHOMBOMERES R2, R4, R5 AND R6 BUT NOT IN R3.
CC EXPRESSION IN R3 IS SEEN AT LATER STAGES AND IS DEPENDENT ON
NEIGHBOURING INTERACTIONS.
CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X87609; CA60915.1; -.
DR EMBL: U34589; AAA92005.1; -.
DR HSP, P00998; ITGS.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00050; kazal. 3.
DR SMART: SM00274; FOLN; 3.
DR SMART: SM00280; KAZAL; 3.
DR PROSITE: PS00282; KAZAL; FALSE-NEG.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 28
FT CHAIN 29 343
FT DOMAIN 93 116 FOLLISTATIN.
FT DOMAIN 117 163 KAZAL-LIKE 1.
FT DOMAIN 166 189 FOLLISTATIN 2.
FT DOMAIN 191 238 KAZAL-LIKE 2.
FT DOMAIN 243 267 FOLLISTATIN 3.
FT DOMAIN 269 315 KAZAL-LIKE 3.
FT DOMAIN 323 329 POLY-GU.
FT DISULFID 117 149 BY SIMILARITY.
FT DISULFID 121 142 BY SIMILARITY.
FT DISULFID 131 163 BY SIMILARITY.
FT DISULFID 191 224 BY SIMILARITY.
FT DISULFID 195 217 BY SIMILARITY.
FT DISULFID 206 238 BY SIMILARITY.
FT DISULFID 269 301 BY SIMILARITY.
FT DISULFID 273 294 BY SIMILARITY.
FT DISULFID 283 315 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 P -> F (IN REF. 2).
FT CONFLICT 177 177 T -> S (IN REF. 3).
FT CONFLICT 183 183 T -> L (IN REF. 2).
FT CONFLICT 202 202 E -> G (IN REF. 2).
FT CONFLICT 228 228 R -> E (IN REF. 3).
SO SEQUENCE 343 AA; 38192 MW; 15A76762560E1AAE CRC64;

Query Match 17.4%; Score 120.5; DB 1; Length 343;
Best Local Similarity 25.6%; Pred. No. 0.00023;
Matches 41; Conservative 12; Mismatches 60; Indels 47; Gaps 8;

OY 1 PNTWVDAGSTERECRDQCAAEKCCINVCGLHSCVAARFOSPAAR----- 49
```

```

Db 153 PELEVOYQKCKKTC--RDVLCPSGSSITCVVDQTNNAVCVTCNRCICEPTSSSEQYLGCNDGI 211
QY 50 -----TTAASGEFVCPQGSDDCDIWD---GQPYCR-C 78
Db 212 TYSSACHLRKATCLLGRSLGIAVEGKCIKAKSCEDIQC--SAGKRC-LMDFKVRGRCSLIC 269
QY 79 RDRC--EKESPTCASDGLTYVNCYDAEACLRGLHLI 116
Db 270 DELCPESKSEAVCASDNTTYPECCAMKEACSGVILEV 309

RESULT 4
FSA_HORSE STANDARD; PRT; 344 AA.
AC 062650;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
GN FST.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Follicle;
RA Sugawara Y., Yamanouchi K.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
CC OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
CC -1- SUBUNIT: MONOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB010829; BAA25699.1; -
CC DR InterPro: IPR003645; FOLN.
CC DR InterPro: IPR002350; KAZAL.
CC Pfam: PF00050; KAZAL; 3.
CC DR SMART: SM00274; FOLN; 3.
CC DR SMART: SM00280; KAZAL; 3.
CC DR PROSITE: PS00282; KAZAL; FALSE_NEG.
CC Glycoprotein; Repeat; Signal.
CC SIGNAL 1 29
CC CHAIN 30 344
CC DOMAIN 94 117 FOLLISTATIN.
CC DOMAIN 118 164 FOLLISTATIN 1.
CC DOMAIN 167 190 KAZAL-LIKE 1.
CC DOMAIN 192 239 FOLLISTATIN 2.
CC DOMAIN 244 268 KAZAL-LIKE 2.
CC DOMAIN 270 316 FOLLISTATIN 3.
CC DOMAIN 316 344 KAZAL-LIKE 3.
CC DISULFID 122 150 BY SIMILARITY.
CC DISULFID 132 164 BY SIMILARITY.
CC DISULFID 192 225 BY SIMILARITY.
CC DISULFID 196 218 BY SIMILARITY.
CC DISULFID 207 239 BY SIMILARITY.
CC DISULFID 270 302 BY SIMILARITY.
CC DISULFID 274 295 BY SIMILARITY.
CC DISULFID 284 316 BY SIMILARITY.
CC CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

SQ SEQUENCE 344 AA; 38002 MW; 18035777D2D9BE4AA CRC64;
Query Match 17.4%; Score 120.5; DB 1; Length 344;
Best Local Similarity 26.5%; Pred. No. 0.00023;
Matches 43; Conservative 15; Mismatches 53; Indels 51; Gaps 10;
QY 1 PNLVWVAQSTGCEECRRDDCAAAEKGCIN-----VCG----- 33
Db 154 PELEVOYQKCKKTC--RDVNCPSGSSITCVVDQTNNAVCVTCNRCICEPTSSSEQYLGCNDGV 212
QY 34 -----LH-----SCVAFPPG--SPAAPTAAASGEFVCPQGSDDCDIWD---GQPYCR-- 77
Db 213 TYSSACHLRKATCLLGRSLGIAVEGKCIKAKSCEDIQC--TGKRC-LMDFKVRGRCSLIC 270
QY 78 ---CRDRCESPTCASDGLTYVNCYDAEACLRGLHLI 116
Db 271 DELCPESKSEEP--VCASDNATYASCCAMKEACSGVILEV 310

RESULT 5
FSA_HUMAN STANDARD; PRT; 338 AA.
AC P19883;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
GN FST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=86247993; PubMed=3380788;
RA Shimazaki S., Koga M., Esch F., Cooksey K., Mercado M., Koba A.,
RA Ueno N., Ying S.Y., Ling N., Guillemin R.;
RL "Primary structure of the human follistatin precursor and its genomic
RL organization.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:4218-4222(1988).
CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
CC OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
CC -1- SUBUNIT: MONOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19481; AAA35851.1; -
CC EMBL: M19480; AAA35851.1; JOINED.
CC PIR: A32141; A32141.
CC PIR: B32141; B32141.
CC HSSP: P00998; ITGS.
CC MIM: 136470; -
CC DR InterPro: IPR003645; FOLN.
CC DR InterPro: IPR002350; KAZAL.
CC Pfam: PF00050; KAZAL; 3.
CC DR SMART: SM00274; FOLN; 3.
CC DR SMART: SM00280; KAZAL; 3.
CC DR PROSITE: PS00282; KAZAL; FALSE_NEG.
CC Glycoprotein; Repeat; Signal; Alternative splicing.
CC SIGNAL 1 29

```

```

FT CHAIN 30 338 FOLLISTATIN.
FT DOMAIN 94 117 FOLLISTATIN 1.
FT DOMAIN 118 164 KAZAL-LIKE 1.
FT DOMAIN 167 190 FOLLISTATIN 2.
FT DOMAIN 192 239 KAZAL-LIKE 2.
FT DOMAIN 244 268 FOLLISTATIN 3.
FT DOMAIN 270 316 KAZAL-LIKE 3.
FT DOMAIN 318 327 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DISULFID 118 150 BY SIMILARITY.
FT DISULFID 122 143 BY SIMILARITY.
FT DISULFID 132 164 BY SIMILARITY.
FT DISULFID 192 225 BY SIMILARITY.
FT DISULFID 196 218 BY SIMILARITY.
FT DISULFID 207 239 BY SIMILARITY.
FT DISULFID 270 302 BY SIMILARITY.
FT DISULFID 274 295 BY SIMILARITY.
FT DISULFID 284 316 BY SIMILARITY.
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 318 338 MISSING (IN ISOFORM 2).
SQ SEQUENCE 338 AA; 37374 MW; 9EAA51B47F02EB30 CRC64;

Query Match 16.88; Score 116.5; DB 1; Length 338;
Best Local Similarity 25.98; Pred. No. 0.00052;
Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8;

OY 1 PNLVMDQSTCERCSDQCAAEKCCINVCGLHSCVAA-RFPGSPAAP----- 49
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 154 PELEVOYQGRCKKTC-RDVECPGSSSTCVVDQTNNAVCTGNRCICPEPASEBOYLGGNDGV 212
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 50 -----TTAASCEGFVCPQGGSDCDIWD--GQPVCR-- 77
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 213 TYSSACHLRATCLGSLGSLAEKCKIKAKSCEDIQC-TGCKKC-LMDKRVGKRCSLC 270
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 78 ---CRDCEKEPSEFTCASDGLTYNNRCYMDAEACLRGLHLHI 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 DELCPDSKSDPE--VCASDNATYASSECAMKEACSSGVLLLEV 310
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
FSL1_HUMAN STANDARD; PRT; 308 AA.
ID FSL1_HUMAN
AC 012841;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN 1 PRECURSOR.
GN FSTL1 OR FRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindaes; Homo.
OX NCB1_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=95045570; PubMed=7957230;
RA Medvise A., Blockx H., van Arnhem W., Willems J., Franssen L.,
RA Devos K., Raymakers J., van de Voorde A., Slegers H.;
RT "Characterization of a rat C6 glioma-secreted follistatin-related
RT protein (FRP). Cloning and sequence of the human homologue.";
RL Eur. J. Biochem. 225:937-946(1994).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE-Synovium;
RX MEDLINE=99000396; PubMed=9786430;
RA Tanaka M., Ozaki S., Osakada F., Mori K., Okubo M., Nakao K.;
RT "Cloning of follistatin-related protein as a novel autoantigen in
RT systemic rheumatic diseases.";
RL Int. Immunol. 10:1305-1314(1998).
RN 3
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney;

```

```

RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
CC PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -! SIMILARITY: CONTAINS 1 FOLLISTATIN-LIKE DOMAIN.
CC -! SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC -! SIMILARITY: CONTAINS 1 WFEC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U06863; AAA6062.1; -.
DR EMBL; D89937; BAA28707.1; -.
DR EMBL; BC000055; AAH00055.1; -.
DR HSSP; P02633; IBOC.
DR MTM; 605547; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR003645; FOLN.
DR InterPro; IPR001239; KAZAL_inhib.
DR InterPro; IPR001007; WFEC.
DR InterPro; IPR002350; KAZAL.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00050; KAZAL; 1.
DR PRINTS; PR00290; KAZAL1_HBTR.
DR SMART; SM00274; FOLN; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00011; VWC_def; 1.
DR PROSITE; PS00282; KAZAL; FALSE_NEG.
DR PROSITE; PS01208; VWC; FALSE_NEG.
KW Glycoprotein; Signal; Heparin-binding.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 308 FOLLISTATIN-RELATED PROTEIN 1.
FT DOMAIN 30 53 FOLLISTATIN.
FT DOMAIN 54 98 KAZAL-LIKE.
FT DOMAIN 233 287 WFEC.
FT DISULFID 54 84 BY SIMILARITY.
FT DISULFID 58 77 BY SIMILARITY.
FT DISULFID 66 98 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 308 AA; 34985 MW; BD4C651FAFF24800 CRC64;

Query Match 16.78; Score 115.5; DB 1; Length 308;
Best Local Similarity 23.58; Pred. No. 0.0006;
Matches 27; Conservative 16; Mismatches 31; Indels 41; Gaps 4;

OY 4 WVDNQSTCERCSDQCAAEKCCINVCGLHSCVAA-RFPGSPAAPTASCGSPVQPQ 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 WVRN-----EEBRSKSKTCANVFC----- 37
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 63 QGSDCDIWD-GQPVCRDRCERKEPSEFTCASDGLTYNNRCYMDAEACLRGLHLHI 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 38 AGRCAVTEKEGPEPCLIEGCKPKRPVCGSNGKTYLNHCHLRHDACTGSKIYV 92
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
FSL1_MACFA STANDARD; PRT; 308 AA.
ID FSL1_MACFA
AC 098X0;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN 1 PRECURSOR.
GN FSTL1 OR OCC1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).

```

```

GN  AGRN.
OC  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
RN  NCBI_TaxID=9031;
RN  [1]
RC  SEQUENCE FROM N.A.
RP  TISSUE=Brain;
RX  MEDLINE=92232297; PubMed=1314620;
RT  Ts'ao K.W.K., Ruegg M.A., Escher G., Kroegeer S., McMahon U.J.;
RT  "cDNA that encodes active agrin.";
RL  Neuron 8:677-689(1992).
RN  [2]
RP  ALTERNATIVE SPLICING.
RX  MEDLINE=92232298; PubMed=1314621;
RT  Ruegg M.A., Ts'ao K.W.K., Horton S.E., Kroegeer S., Escher G.,
RT  Gensch E.M., McMahon U.J.;
RT  "The agrin gene codes for a family of basal lamina proteins that
RT  differ in function and distribution.";
RL  Neuron 8:691-699(1992).
CC  -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
CC  AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
CC  ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
CC  -1- SUBUNIT: BINDS TO LAMININ.
CC  -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
CC  JUNCTION.
CC  -1- ALTERNATIVE PRODUCTS: AT LEAST THREE DIFFERENT FORMS ARISE BY
CC  ALTERNATIVE SPLICING, THEY DIFFER IN THEIR ACETYLCHOLINE RECEPTOR
CC  CLUSTERING ACTIVITY.
CC  -1- PTM: CONTAINS HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-
CC  LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC  -1- SIMILARITY: CONTAINS 9 KAZAL-LIKE DOMAINS.
CC  -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC  -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC  -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL: M94271; AAA48585.1; -.
DR  EMBL: M97371; AAA48586.1; -.
DR  EMBL: M97372; -. NOT_ANNOTATED_CDS.
DR  PIR: JH0591; AGCH.
DR  InterPro: IPR000152; ASX_hydroxyl.
DR  InterPro: IPR000561; EGF-like.
DR  InterPro: IPR000742; EGF_2.
DR  InterPro: IPR003884; FACI_MAC.
DR  InterPro: IPR003645; FOIN.
DR  InterPro: IPR001239; kazal_inhib.
DR  InterPro: IPR002049; laminin_EGF.
DR  InterPro: IPR001791; laminin_G.
DR  InterPro: IPR000082; SEA.
DR  InterPro: IPR002350; kazal.
DR  Pfam: PF00008; EGF_4.
DR  Pfam: PF00050; kazal; 9.
DR  Pfam: PF00053; laminin_EGF_2.
DR  Pfam: PF00054; laminin_G; 3.
DR  Pfam: PF01390; SEA; 1.
DR  PRINTS: PR00290; KAZALINHBTR.
DR  SMART: SM0180; EGF_Lam; 2.
DR  SMART: SM00001; EGF_like; 4.
DR  SMART: SM00057; FIMAC; 2.
DR  SMART: SM00274; FOIN; 5.
DR  SMART: SM00280; KAZAL; 9.
DR  SMART: SM00282; LamG; 3.
DR  SMART: SM00200; SEA; 1.
DR  PROSITE: PS00010; ASX_HYDROXYL; 1.

```

DR PROSITE: PS00022; EGF_1; 6.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF_1.
 DR PROSITE: PS50024; SEA; 1.
 KW Lamprin EGF-like domain; Repeat; Alternative splicing; Signal;
 EFT CHAIN 1 38
 EFT SIGNAL 39 1955
 EFT CHAIN 54 126
 EFT DOMAIN 130 201
 EFT DOMAIN 202 273
 EFT DOMAIN 276 344
 EFT DOMAIN 350 418
 EFT DOMAIN 419 483
 EFT DOMAIN 484 548
 EFT DOMAIN 551 633
 EFT DOMAIN 675 728
 EFT DOMAIN 729 775
 EFT DOMAIN 781 851
 EFT DOMAIN 856 995
 EFT DOMAIN 1150 1219
 EFT DOMAIN 1229 1265
 EFT DOMAIN 1446 1483
 EFT DOMAIN 1485 1522
 EFT DOMAIN 1714 1752
 EFT DOMAIN 86 105
 EFT DISULFID 94 126
 EFT DISULFID 160 180
 EFT DISULFID 169 201
 EFT DISULFID 233 252
 EFT DISULFID 241 273
 EFT DISULFID 304 323
 EFT DISULFID 312 344
 EFT DISULFID 378 397
 EFT DISULFID 386 418
 EFT DISULFID 443 462
 EFT DISULFID 451 483
 EFT DISULFID 507 527
 EFT DISULFID 516 548
 EFT DISULFID 592 612
 EFT DISULFID 601 633
 EFT DISULFID 677 694
 EFT DISULFID 696 705
 EFT DISULFID 708 726
 EFT DISULFID 729 741
 EFT DISULFID 731 748
 EFT DISULFID 750 759
 EFT DISULFID 762 773
 EFT DISULFID 810 830
 EFT DISULFID 819 851
 EFT DISULFID 1233 1244
 EFT DISULFID 1238 1253
 EFT DISULFID 1255 1264
 EFT DISULFID 1450 1461
 EFT DISULFID 1455 1471
 EFT DISULFID 1473 1482
 EFT DISULFID 1489 1500
 EFT DISULFID 1494 1510
 EFT DISULFID 1512 1521
 EFT DISULFID 1718 1731
 EFT DISULFID 1725 1740
 EFT DISULFID 1742 1751
 EFT CARBOHYD 390 390
 EFT CARBOHYD 659 659
 EFT CARBOHYD 764 764
 EFT CARBOHYD 814 814
 EFT VARSPLIC 1648 1651
 EFT VARSPLIC 1783 1793
 EFT CONFLICT 1129 1131
 EFT SEQUENCE 1955 AA; 211411 MW; BADB27C2342581 CRC64;

Query Match 16.7%; Score 115.5; DB 1; Length 1955;
 Best Local Similarity 27.1%; Pred. No. 0.0026; Mismatches 50; Indels 29; Gaps 7;
 Matches 36; Conservative 18; Mismatches 50; Indels 29; Gaps 7;
 QY 7 AOSTGER-----ECSPD-----ODCAAEKCCINVCGL---HSCVAFPPGSPAP 49
 Db 295 ARCSCHDILTDGTYRYPVCAKRSRYNSDCEKQKCHOKAIPVKHS-----GPCDL 346
 QY 50 TPAASCEGVPCPQGGSDIDWDGPPVCRDRCKRDEPSFTCASGLTYNRCYMDAEC 109
 Db 347 GTPSPCSEVCE-TFGATCVKKNREPVCQVCGQGRYDPVCGSDNRYGNPCELNMAVCV 405
 QY 110 --RGHL-HIYPC 119
 Db 406 LKREIRVHKHGPC 418
 RESULT 9
 FSA_SHEEP STANDARD: PRT; 337 AA.
 AC P31514;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN) (FRAGMENT).
 OS EST.
 GN Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_Taxid:9940;
 RN [1]
 RP MEDLINE-92337809; Pubmed-1632897;
 RX Tisdall D.J., Hill D., Petersen G.B., Fleming J.S.;
 RA "ovine follistatin: characterization of cDNA and expression in sheep
 RT ovary during the luteal phase of the oestrous cycle.";
 RL J. Mol. Endocrinol. 8:259-264(1992).
 CC - FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 CC OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
 CC - SUBUNIT: MONOMER (POTENTIAL).
 CC - SUBCELLULAR LOCATION: SECRETED.
 CC - SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
 CC - SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M63123; AAA31522.1; -.
 DR HSSP: P00996; ITGS.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR002350; kazal.
 DR Pfam: PF00050; kazal; 3.
 DR SMART: SM00274; FOLN; 3.
 DR SMART: SM00280; KAZAL; 3.
 DR PROSITE: PS00282; KAZAL; FALSE-NEG.
 KW Glycoprotein; Repeat; Signal.
 EFT CHAIN 1 22
 EFT SIGNAL 23 337
 EFT CHAIN 87 110
 EFT DOMAIN 111 157
 EFT DOMAIN 160 183
 EFT DOMAIN 185 232
 EFT DOMAIN 237 261
 EFT DOMAIN 263 309
 EFT FOLLISTATIN 1.
 EFT KAZAL-LIKE 1.
 EFT FOLLISTATIN 2.
 EFT KAZAL-LIKE 2.
 EFT FOLLISTATIN 3.
 EFT KAZAL-LIKE 3.

```

FT DOMAIN 314 326 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DISULFID 111 143 BY SIMILARITY.
FT DISULFID 115 136 BY SIMILARITY.
FT DISULFID 125 157 BY SIMILARITY.
FT DISULFID 165 218 BY SIMILARITY.
FT DISULFID 189 211 BY SIMILARITY.
FT DISULFID 200 232 BY SIMILARITY.
FT DISULFID 263 295 BY SIMILARITY.
FT DISULFID 267 288 BY SIMILARITY.
FT DISULFID 277 309 BY SIMILARITY.
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 337 AA; 37082 MW; 1EBBEBB6B109C4 CRC64;

Query Match 16.4%; Score 113.5; DB 1; Length 337;
Best Local Similarity 26.5%; Pred. No. 0.00098;
Matches 43; Conservative 15; Mismatches 33; Indels 51; Gaps 11;

QY 1 PNLWVDAQSTCERCSRDQCAAEKCCIN-----VCG----- 33
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 147 PELLEVQYQCKKTC-RDVEFCGSSSTCVVDQTNNAVCYTCNRIKCEPTSSSEQYLGGNDGV 205
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 34 -----LH-----SCVAARFPG--SPAAPTAAACGEGFVCPQGGSDCDIWD---GQPYCR-C 78
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 206 TTPSACHLRKATCLGRISGLAYEGKCIKAKSCEDIQC-TGGKRC-LMDFKVGRGRCSLC 263
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 79 RDRC---EKEPFTCASDGLTYNRCYMDAECRLGHLHT 116
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 264 GELCPESKSEEP--VCASDNATYASECAMKEAKCASSGVILEV 303
   | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
FSA_MOUSE STANDARD; PRT; 344 AA.
ID FSA_MOUSE
AC P47931;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
GN FST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95324368; PubMed=7600958;
RA Albano R.M., Atkell R., Beddington R.S.P., Smith J.C.;
RT "Expression of inhibin subunits and follistatin during
RT postimplantation mouse development: decidual expression of activin
RT and expression of follistatin in primitive streak, somites and
RT hindbrain.";
RL Development 120:803-813(1994).
RN [2]
RP SEQUENCE OF 3-340 FROM N.A.
RC STRAIN=CBA X NMRI; TISSUE=Ovary;
RX MEDLINE=95045246; PubMed=7956942;
RA Tuuri T., Eramaa M., Hildén K., Rittvos O.;
RT "Activin-binding protein follistatin messenger ribonucleic acid and
RT secreted protein levels are induced by chorionic gonadotropin in
RT cultured human granulosa-luteal cells.";
RL Endocrinology 135:2196-2203(1994).
CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
CC OF PITUITARY FOLLICLE STIMULATING HORMONE (FSH).
CC -1- SUBUNIT: MONOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DEVELOPMENTAL STAGE: EMBRYONIC EXPRESSION FIRST OCCURS IN THE
CC PRIMITIVE STREAK, FOLLOWED BY EXPRESSION IN HEAD MESODERM,
CC SOMITES, AND SPECIFIC RHOMBOMERES OF THE HINDBRAIN, AND LATER IN
CC MIDBRAIN AND DIENCEPHALON. NO EXPRESSION IS SEEN IN THE NODE OR

```

```

CC NOTOCHORD.
CC -1- SIMILARITY: CONTAINS 3 FOLLISTATIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL: 229532; CAA82648.1; -.
CC EMBL: X83377; CAA58291.1; -.
DR HSSP; P00998; ITGS.
DR MGD; MGI:95586; FSL.
DR InterPro; IPR003645; FOLIN.
DR InterPro; IPR002350; kazal.
DR Pfam; PF00050; kazal; 3.
DR SMART; SM00274; FOLN; 3.
DR SMART; SM00280; KAZAL; 3.
DR PROSITE; PS00282; KAZAL; FALSE_NEG.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29
FT CHAIN 30 344 FOLLISTATIN.
FT DOMAIN 94 117 FOLLISTATIN 1.
FT DOMAIN 118 164 KAZAL-LIKE 1.
FT DOMAIN 167 190 FOLLISTATIN 2.
FT DOMAIN 192 239 KAZAL-LIKE 2.
FT DOMAIN 244 268 FOLLISTATIN 3.
FT DOMAIN 270 316 KAZAL-LIKE 3.
FT DISULFID 118 150 BY SIMILARITY.
FT DISULFID 122 143 BY SIMILARITY.
FT DISULFID 132 164 BY SIMILARITY.
FT DISULFID 192 225 BY SIMILARITY.
FT DISULFID 196 218 BY SIMILARITY.
FT DISULFID 207 239 BY SIMILARITY.
FT DISULFID 270 302 BY SIMILARITY.
FT DISULFID 274 295 BY SIMILARITY.
FT DISULFID 284 316 BY SIMILARITY.
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 241 242 KA -> T (IN REF. 1).
SQ SEQUENCE 344 AA; 37866 MW; 935B6CBB213176P9 CRC64;

Query Match 16.4%; Score 113.5; DB 1; Length 344;
Best Local Similarity 26.5%; Pred. No. 0.001;
Matches 43; Conservative 13; Mismatches 55; Indels 51; Gaps 10;

QY 1 PNLWVDAQSTCERCSRDQCAAEKCCIN-----VCG----- 33
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 154 PELLEVQYQCKKTC-RDVEFCGSSSTCVVDQTNNAVCYTCNRIKCEPTSSSEQYLGGNDGV 212
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 34 -----LH-----SCVAARFPG--SPAAPTAAACGEGFVCPQGGSDCDIWD---GQPYCR-- 77
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 213 TTPSACHLRKATCLGRISGLAYEGKCIKAKSCEDIQC-GGKRC-LMDSKVGRGRCSLC 270
   | | | | | | | | | | | | | | | | | | | | | | | |
QY 78 -----CRDCEKEPFTCASDGLTYNRCYMDAECRLGHLHT 116
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 271 DELCPDSKSDP--VCASDNATYASECAMKEAKCASSGVILEV 310
   | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
FSA_PIG STANDARD; PRT; 344 AA.
ID FSA_PIG
AC P10669; P10670;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLISTATIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
GN FST.
OS Sus scrofa (Pig).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OC NCBI_TaxId=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88209050; Pubmed=3365249;
 RA Shimasaki S., Koga M., Esch F., Mercado M., Cooksey K., Koba A.,
 RT ling N.;
 RT "Follicle-stimulating gene structure supports two forms of mature
 RT follicle-stimulating hormone produced by alternative splicing."
 RT Biochem. Biophys. Res. Commun. 152:717-723(1988).
 RL Mol. Endocrinol. 1:849-855(1987).
 RN (2)
 RP SEQUENCE OF 1-334 FROM N.A.
 RX MEDLINE=91042571; Pubmed=3153465;
 RA Esch F.S., Shimasaki S., Mercado M., Cooksey K., Ling N., Ying S.,
 RA Ueno N., Guillemin R.;
 RT "Structural characterization of follicle-stimulating: a novel follicle-
 RT stimulating hormone release-inhibiting polypeptide from the gonad.";
 RT Mol. Endocrinol. 1:849-855(1987).
 CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 CC OF PITUITARY FOLLICLE-STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND A'; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 3 FOLLICULIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M19529; AAA31036.1; -
 DR EMBL: M19529; AAA31037.1; -
 DR EMBL: M36512; AAA31038.1; -
 DR PIR: A27701; A27701.
 DR HSSP: P00998; ITGS.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR002350; kazal.
 DR Pfam: PF00050; kazal; 3.
 DR SMART: SM00274; FOLN; 3.
 DR SMART: SM00280; KAZAL; 3.
 DR PROSITE: PS00282; KAZAL; FALSE_NEG.
 DR GlycoProtein: Repeat; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 344
 FT DOMAIN 94 117 FOLLICULIN 1.
 FT DOMAIN 118 164 KAZAL-LIKE 1.
 FT DOMAIN 167 190 FOLLICULIN 2.
 FT DOMAIN 192 239 KAZAL-LIKE 2.
 FT DOMAIN 244 268 FOLLICULIN 3.
 FT DOMAIN 270 316 KAZAL-LIKE 3.
 FT DOMAIN 321 333 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 334 344 BY SIMILARITY.
 FT DISULFID 118 150 BY SIMILARITY.
 FT DISULFID 122 143 BY SIMILARITY.
 FT DISULFID 132 164 BY SIMILARITY.
 FT DISULFID 192 225 BY SIMILARITY.
 FT DISULFID 196 218 BY SIMILARITY.
 FT DISULFID 207 239 BY SIMILARITY.
 FT DISULFID 270 302 BY SIMILARITY.
 FT DISULFID 274 295 BY SIMILARITY.
 FT DISULFID 284 316 BY SIMILARITY.
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 318 344 MISSING (IN ISOFORM A').
 FT SEQUENCE 344 AA; 38035 MM; 6906E7A23CF9BD21 CRC64;

Query Match 16.4%; Score 113.5; DB 1; Length 344;
 Best Local Similarity 26.5%; Pred. No. 0.001;
 Matches 43; Conservative 15; Mismatches 53; Indels 51; Gaps 11;
 QY 1 PNLVVAQSTCERCSRDQCAAEKCCIN-----VCG-----33
 Db 154 PELEVOYQKCKKRC-RDVCPSGSSCTCVNATNAYCTCNRICPEPTSSRQYLCGNDGV 212
 QY 34 -----LI-----SCVAARPPG--SPAAPTAAACGCEYVCPDQSDCDIWD---GQPVCR-C 78
 Db 213 TYSSACHLRATKATCLGRSIGLAEGKCIKAKSCEDIQC-TGKKKC-LMDFKVRGRCSLTC 270
 QY 79 RDRC-----EKESFTCASDGLTYNRCYMAEAFLRLHLHI 116
 Db 271 DELCPESKSEEP--VCASDNATYASCEAMKEACSSGVLEEV 310
 RESULT 12
 FSA_RAT STANDARD; PRT; 344 AA.
 ID FSA_RAT
 AC P21674;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FOLLICULIN PRECURSOR (FS) (ACTIVIN-BINDING PROTEIN).
 GN EST.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxId=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE=89261821; Pubmed=2725528;
 RA Shimasaki S., Koga M., Buscaglia M.L., Simmons D.M., Blozak T.A.,
 RA Ling N.;
 RT "Follicle-stimulating gene expression in the ovary and extragonadal tissues."
 RT Mol. Endocrinol. 3:651-659(1989).
 CC -1- FUNCTION: BINDS DIRECTLY TO ACTIVIN AND FUNCTIONS AS AN ACTIVIN
 CC ANTAGONIST. SPECIFIC INHIBITOR OF THE BIOSYNTHESIS AND SECRETION
 CC OF PITUITARY FOLLICLE-STIMULATING HORMONE (FSH).
 CC -1- SUBUNIT: MONOMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: CONTAINS 3 FOLLICULIN-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M31591; AAB60704.1; -
 DR EMBL: M31586; AAB60704.1; JOINED.
 DR EMBL: M31587; AAB60704.1; JOINED.
 DR EMBL: M31588; AAB60704.1; JOINED.
 DR EMBL: M31589; AAB60704.1; JOINED.
 DR EMBL: M31590; AAB60704.1; JOINED.
 DR HSSP: P00998; ITGS.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR002350; kazal.
 DR Pfam: PF00050; kazal; 3.
 DR SMART: SM00274; FOLN; 3.
 DR SMART: SM00280; KAZAL; 3.
 DR PROSITE: PS00282; KAZAL; FALSE_NEG.
 DR GlycoProtein: Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 344
 FT DOMAIN 94 117 FOLLICULIN 1.
 FT DOMAIN 118 164 KAZAL-LIKE 1.
 FT DOMAIN 167 190 FOLLICULIN 2.


```

FT DOMAIN 192 239 KATAL-LIKE 2.
FT DOMAIN 244 268 FOLLISTATIN 3.
FT DOMAIN 270 316 KATAL-LIKE 3.
FT DOMAIN 321 333 ASP/GRU-RICH 3.
FT DISULFID 118 150 BY SIMILARITY.
FT DISULFID 122 143 BY SIMILARITY.
FT DISULFID 132 164 BY SIMILARITY.
FT DISULFID 192 225 BY SIMILARITY.
FT DISULFID 196 218 BY SIMILARITY.
FT DISULFID 207 239 BY SIMILARITY.
FT DISULFID 270 302 BY SIMILARITY.
FT DISULFID 274 295 BY SIMILARITY.
FT DISULFID 284 316 BY SIMILARITY.
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 344 AA; 37838 MW; 864244CF05436552 CRC64;

```

```

Query Match 16.4%; Score 113.5; DB 1; Length 344;
Best Local Similarity 27.2%; Pred. No. 0.001;
Matches 44; Conservative 13; Mismatches 54; Indels 51; Gaps 11;

```

```

QY 1 PNLVDAQSTCEKRECSRDDCAAEKC-----CI-----NVCG----- 33
Db 154 PLEFVOYQKCKKTC-RDVPFGSSSTCVVDQNNAYVCNRCIPSPSSSGLCNDGV 212
QY 34 -----LH-----SCVAPRPG--SPAAPTAAGCEGFVCPQSGDCDIND--GQVPCR-- 77
Db 213 TVSSACHLRKATCLLRSLGSLAYEGKCIKAKSCEDIQG-GGKKC-LMFVKYGRGCSLC 270
QY 78 ---CRDCKEPEPFCASGLTYNNCYMDARCLRGLHLI 116
Db 271 DELCPDSKDEP--VCASDNATYASECAKREAKACSSGVLEV 310

```

```

RESULT 13
AGRI_RAT STANDARD: PRT: 1959 AA.
AC P25304; 063034;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AGRIN PRECURSOR.
GN AGRIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN 11 SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RP TISSUE-Embryonic spinal cord;
RX MEDLINE-9122570; PubMed-1851019;
RA Rupp F., Payan D.G., Magill-Sole C., Cowan D.M., Scheller R.H.;
RT "Structure and expression of a rat agrin.";
RN Neuron 6:811-823(1991).
RP MEDLINE-92407628; PubMed-1326608;
RA Rupp F., Oezcelik T., Linal M., Peterson K., Francke U., Scheller R.;
RT "Structure and chromosomal localization of the mammalian agrin gene.";
RN J. Neurosci. 12:3535-3544(1992).
CC -1- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
CC -1- SUBUNIT: BINDS TO LAMININ.
CC -1- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
CC JUNCTION.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS, 1 (SHOWN HERE), 2, 3, 4
CC AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
CC ACETYLCHOLINE RECEPTOR CLUSTERING ACTIVITY.
CC -1- TISSUE SPECIFICITY: EMBRYONIC NERVOUS SYSTEM AND MUSCLE.
CC -1- DEVELOPMENTAL STAGE: MORE ABUNDANT EARLY IN DEVELOPMENT.
CC -1- PTM: CONTAINS HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-

```

```

CC LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 9 KATAL-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-18 OR MET-24 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isdb-sdb.ch/announce/
CC or send an email to license@isdb-sdb.ch).
CC -----
DR EMBL; M64780; AAA40703.1; -
DR EMBL; M64780; AAA40702.1; ALT_INIT.
DR PIR; JH0399; AGRI.
DR HSP; P00740; IIXA.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003884; FACI_MAC.
DR InterPro; IPR003645; FOIN.
DR InterPro; IPR001239; Katal_inhib.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001921; Laminin_G.
DR InterPro; IPR000082; SEA.
DR Pfam; PF00050; Katal_9.
DR Pfam; PF00053; Laminin_EGF_2.
DR Pfam; PF00054; Laminin_G_3.
DR Pfam; PF01390; SEA_1.
DR PRINTS; PR00290; KATALINHTR.
DR SMART; SM00180; EGF_Lam_2.
DR SMART; SM00001; EGF-like_4.
DR SMART; SM00057; FIMAC_3.
DR SMART; SM00274; FOIN_5.
DR SMART; SM00280; KATAL_9.
DR SMART; SM00282; LamG_3.
DR SMART; SM0200; SEA_1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF_1.
DR PROSITE; PS50024; SEA_1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
KW SIGNAL
FT CHAIN 1 29
FT CHAIN 30 1959
FT DOMAIN 65 137 KATAL-LIKE 1.
FT DOMAIN 141 212 KATAL-LIKE 2.
FT DOMAIN 213 284 KATAL-LIKE 3.
FT DOMAIN 287 356 KATAL-LIKE 4.
FT DOMAIN 361 429 KATAL-LIKE 5.
FT DOMAIN 430 494 KATAL-LIKE 6.
FT DOMAIN 495 559 KATAL-LIKE 7.
FT DOMAIN 563 645 KATAL-LIKE 8.
FT DOMAIN 688 741 LAMININ EGF-LIKE 1.
FT DOMAIN 742 788 LAMININ EGF-LIKE 2.
FT DOMAIN 794 864 SEA.
FT DOMAIN 1023 1145 KATAL-LIKE 9.
FT DOMAIN 1220 1258 EGF-LIKE 1.
FT DOMAIN 1440 1477 EGF-LIKE 2.
FT DOMAIN 1479 1516 EGF-LIKE 3.
FT DOMAIN 1709 1748 EGF-LIKE 4.
FT DOMAIN 869 992 SER/THR-RICH.
FT DOMAIN 1147 1215 SER/THR-RICH.
FT DISULFID 97 116 POTENTIAL.
FT DISULFID 105 137 POTENTIAL.
FT DISULFID 171 191 POTENTIAL.
FT DISULFID 180 212 POTENTIAL.
FT DISULFID 244 263 POTENTIAL.
FT DISULFID 252 284 POTENTIAL.

```



```
FT DISULFID 316 335 POTENTIAL.
FT DISULFID 324 356 POTENTIAL.
FT DISULFID 389 408 POTENTIAL.
FT DISULFID 397 429 POTENTIAL.
FT DISULFID 454 473 POTENTIAL.
FT DISULFID 462 494 POTENTIAL.
FT DISULFID 518 538 POTENTIAL.
FT DISULFID 527 559 POTENTIAL.
FT DISULFID 604 624 POTENTIAL.
FT DISULFID 613 645 POTENTIAL.
FT DISULFID 688 700 BY SIMILARITY.
FT DISULFID 690 707 BY SIMILARITY.
FT DISULFID 709 718 BY SIMILARITY.
FT DISULFID 721 739 BY SIMILARITY.
FT DISULFID 742 754 BY SIMILARITY.
FT DISULFID 744 761 BY SIMILARITY.
FT DISULFID 763 772 BY SIMILARITY.
FT DISULFID 775 786 BY SIMILARITY.
FT DISULFID 823 843 POTENTIAL.
FT DISULFID 832 864 POTENTIAL.
FT DISULFID 1224 1235 BY SIMILARITY.
FT DISULFID 1229 1246 BY SIMILARITY.
FT DISULFID 1248 1257 BY SIMILARITY.
FT DISULFID 1444 1455 POTENTIAL.
FT DISULFID 1449 1465 POTENTIAL.
FT DISULFID 1467 1476 POTENTIAL.
FT DISULFID 1483 1494 BY SIMILARITY.
FT DISULFID 1488 1504 BY SIMILARITY.
FT DISULFID 1506 1515 BY SIMILARITY.
FT DISULFID 1713 1727 BY SIMILARITY.
FT DISULFID 1721 1736 BY SIMILARITY.
FT DISULFID 1738 1747 BY SIMILARITY.
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1144 1152 MISSING (IN ISOFORM 2).
FT VARSPLIC 1780 1798 MISSING (IN ISOFORM 3).
FT VARSPLIC 1780 1798 MISSING (IN ISOFORM 4).
FT VARSPLIC 1780 1787 MISSING (IN ISOFORM 5).
FT VARIANT 314 314 V -> VTCD (IN A VARIANT).
SQ SEQUENCE 1959 AA; 208645 MW; 7EEFDEFAFF89C31 CRC64;
```

Query Match 16.2%; Score 112; DB 1; Length 1959;

Best Local Similarity 28.9%; Pred. No. 0.0055; Matches 35; Conservative 14; Mismatches 48; Indels 24; Gaps 6;

```
QY 7 AOSTERECES--RDQCAAEKCCINVCGLHS--C-----VAARPPGSPAAPTAAASC 55
DB 381 AECECORVCSGIYDPVCGSGVYGVSCLESMACITLGRITQVARRGCPD-----CGQC 435
QY 56 EGEFVQSGSDCDIMDPQVRCRCRCEKPEPFTCASDGLTYNRCYMDAEACIRGLHLH 115
DB 436 -----RFGSLCEVETGR--CVCPSCEVESAGPVCGSDGHTYASSECLHYNACTHQLSTLY 487
QY 116 I 116
DB 488 V 488
```

RESULT 14
FSL_MOUSE STANDARD; PRT; 306 AA.
AC 062356; 099J19;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FOLLICULIN-RELATED PROTEIN 1 PRECURSOR (TGF-BETA-INDUCIBLE PROTEIN
TSC-36).
GN FSTLI OR FSTL OR FRP OR TSC36.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039028; Pubmed=7901004;
RA Shibamura M., Mashimo J., Mita A., Kuroki T., Nose K.;
RT "Cloning from a mouse osteoblastic cell line of a set of
RT transforming-growth-factor-beta 1-regulated genes, one of which
RT seems to encode a follistatin-related polypeptide."
RL Eur. J. Biochem. 217:13-19(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY MODULATE THE ACTION OF SOME GROWTH FACTORS ON CELL
CC PROLIFERATION AND DIFFERENTIATION. BINDS HEPARIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 FOLLISTATIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M91380; AAC37633.1; -
CC EMBL: BC006185; AAH06185.1; -
CC HSSP: P01001; 1BUS.
DR MGD; MGI:102793; PstI.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR003645; FOLN.
DR InterPro; IPR001239; kazal_inh1b.
DR InterPro; IPR002350; kazal.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00050; kazal; 1.
DR PRINTS; PRO0290; KAZALINHBTR.
DR SMART; SM00274; FOLN; 1.
DR SMART; SM00280; KAZAL; 1.
DR PROSITE; PS00282; KAZAL; FALSE NEG.
DR PROSITE; PS01208; VMFC; FALSE NEG.
KM Glycoprotein; Signal; Heparin-binding.
FT SIGNAL 1 18 BY SIMILARITY.
FT CHAIN 19 306 FOLLISTATIN-RELATED PROTEIN 1.
FT DOMAIN 28 51 FOLLISTATIN.
FT DOMAIN 52 96 KAZAL-LIKE.
FT DOMAIN 231 285 VMFC.
FT DISULFID 52 82 BY SIMILARITY.
FT DISULFID 56 75 BY SIMILARITY.
FT DISULFID 64 96 BY SIMILARITY.
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 235 235 V -> D (IN REF. 2).
SQ SEQUENCE 306 AA; 34538 MW; 4631070AE81EBFC4 CRC64;
```

Query Match 16.1%; Score 111.5; DB 1; Length 306;

Best Local Similarity 23.8%; Pred. No. 0.0014; Matches 24; Conservative 16; Mismatches 30; Indels 31; Gaps 3;

```
QY 18 DQCAAEKCCINVCGLHSCVARRPPGSPAAPTAAASCEGFVQSGSDCDIMD-GQPV 75
DB 19 EEPERSKSKICANVFEC-----AGRECAVTEKEPPT 49
QY 76 CCRDRCEKEPFTCASDGLTYNRCYMDAEACIRGLHLHI 116
DB 50 CLCIQCKPKHRRPVCGSGNGKTYLNHCLHRDACITGSKIQV 90
```


This Page Blank (uspto)

A:Experimental source: testis
A:Accession: B32141
A:Molecule type: mRNA
A:Residues: 1317 <SH2>
A:Cross-references: GB:M19481; GB:J03771; NID:9182719; PID:AAA35851.1; PID:9182721
A:Experimental source: testis
C:Superfamily: follistatin; Kazal proteinase inhibitor homology
C:Keywords: alternative splicing; duplication; gonad
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-344/Product: follistatin 1 #status predicted <MAT1>
F:30-317/Product: follistatin 2 #status predicted <MAT2>
F:265-316/Domain: Kazal proteinase inhibitor homology <KPI6>

Query Match 16.8%; Score 116.5; DB 2; Length 344;
Best Local Similarity 25.9%; Pred. No. 0.0077;
Matches 42; Conservative 11; Mismatches 58; Indels 51; Gaps 8;

1 PNLVDAOSTCERECRSHDQCAAEKCCINVCGLHSCVAA-RFPGSPAP----- 49
154 PELEVQYQGRCKKTC-RDVECPGSGTCVVDPTNNAYCYTCNRICPEPASEQYLQGDGV 212
50 -----TTAASCEGFVCPQGGSDCDIWD--GQPVCR-- 77
213 TYSACHLRKATCCLGSLAYEGKCIKAKSCEDIQ-TGKKC-LMDFVGRGCSLC 270
78 ----CRDCEKEPFTCASDGLTYNRCYMDAECRLHLHI 116
271 DELCPDSKSDP--VCASDMNTVASECAMKEAACSGVLLLEV 310

RESULT 7
Follistatin-related protein - human
C:Species: Homo sapiens (hmn)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 17-Nov-2000
C:Accession: S51362
R:Wijsten, A.; Blockx, H.; van Arnhem, W.; Willems, J.; Franssen, L.; Devos, K.; Raynacke
Eur. J. Biochem. 225, 937-946, 1994
A:Title: Characterization of a rat C(6) glioma-secreted follistatin-related protein (FRL
A:Reference number: S51361; MUID:95045570
A:Accession: S51362
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-308 <ZM1>
A:Cross-references: EMBL:U06863; NID:9536897; PID:AAA6062.1; PID:9536898
A:Genetics:
A:Gene: GDB:FRP
A:Cross-references: GDB:434868
C:Superfamily: Kazal proteinase inhibitor homology
F:50-98/Domain: Kazal proteinase inhibitor homology <KPI5>

Query Match 16.7%; Score 115.5; DB 2; Length 308;
Best Local Similarity 23.5%; Pred. No. 0.0087;
Matches 27; Conservative 16; Mismatches 31; Indels 41; Gaps 4;

4 WVDASTCERECRSHDQCAAEKCCINVCGLHSCVAA-RFPGSPAPTTAASCEGFVCPQ 62
17 WVA-----EELRSKSKICANVFC----- 37
63 QGSDCDIWD-GQPVCRDRCERKEPFTCASDGLTYNRCYMDAECRLHLHI 116
38 AGRECAVTEKGEPTCLTEOCKPHKRPVCGSNGKTYLNHCLHRDACLTGSKIQV 92

RESULT 8
AGCH
agrin precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C:Accession: JH0591; A38857; I50692
R:Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroegeer, S.; McManhan, U.J.

Neuron 8, 677-689, 1992
A:Title: cDNA that encodes active agrin.
A:Reference number: JH0591; MUID:92232297
A:Accession: JH0591
A:Molecule type: mRNA
A:Residues: 1-1955 <TS1>
A:Cross-references: GB:M94271; NID:9211120; PID:AAA48585.1; PID:9211121
A:Experimental source: brain
R:Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroegeer, S.; Escher, G.; Gensch, E.M.; Mc
Neuron 8, 691-699, 1992
A:Title: The agrin gene codes for a family of basal lamina proteins that differ in f
A:Reference number: A38857; MUID:92232298
A:Contents: alternative splicing
A:Accession: A38857
A:Molecule type: mRNA
A:Residues: 1132-1783; 1795-1955 <RU2>
A:Cross-references: GB:M97371
A:Accession: B38857
A:Molecule type: mRNA
A:Residues: 1221-1647; 1652-1783; 1794-1955 <RU3>
A:Cross-references: GB:M97372
A:Note: translation of the nucleotide sequence is not complete
R:Thomae, W.S.; O'Dowd, D.K.; Smith, M.A..
Dev. Biol. 158, 523-535, 1993
A:Title: Developmental expression and alternative splicing of chick agrin RNA.
A:Reference number: I50692; MUID:93345745
A:Accession: I50692
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'SHLSNEIPA', 1784-1795 <THO>
A:Cross-references: EMBL:U07271; NID:9459665; PID:AAA16788.1; PID:9459666
C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholin
C:Comment: Alternative splicing produces two inactive proteins: agrin-related protein
C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G r
C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-1955/Product: agrin #status predicted <MAT>
F:39-1783; 1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
F:39-1647; 1652-1783; 1794-1955/Product: agrin-related protein 2 #status predicted <AG
F:77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F:152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F:225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F:295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
F:370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F:435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F:500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F:584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F:675-726/Domain: laminin-type EGF-like homology <LE1>
F:729-773/Domain: laminin-type EGF-like homology <LE2>
F:801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
F:856-995/Region: serine/threonine-rich
F:1150-1219/Region: EGF homology <EG1>
F:1233-1264/Domain: EGF homology <EG2>
F:1429-1431/Region: motor neuron attachment (L-R-E) motif
F:1489-1482/Domain: EGF homology <EG3>
F:1560-1711/Domain: laminin G repeat homology <LG2>
F:1718-1751/Domain: EGF homology <EG4>
F:1803-1955/Domain: laminin G repeat homology <LG3>
F:86-105; 94-126; 160-180; 169-201; 233-252; 241-273; 304-323; 312-344; 378-397; 386-418; 443-
1482; 1489-1500; 1494-1510; 1512-1521/Disulfide bonds: #status predicted
F:390; 659; 764; 814/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 16.7%; Score 115.5; DB 1; Length 1955;
Best Local Similarity 27.1%; Pred. No. 0.032;
Matches 36; Conservative 18; Mismatches 50; Indels 29; Gaps 7;

7 AOSTCER-----ECSDR-----ODCAAEKCCINVCGL--HSCVAARFPGSPAP 49
295 ARCGCDRTTCGTYRPPCARSRFYSNDCEQKAKCHQKAIPVKHS-----GPDLL 346

A:Residues: 1-344 <ESCI>
A:Molecule type: mRNA
A:Cross-references: GB:M36512; GB:M36513; NID:q16446461
A:Accession: BA0064
A:Status: preliminary
A:Molecule type: mRNA

A/Accession: 15/6/6
A/Status: preliminary; translated from GB/
A/Molecule type: DNA
A/Residues: 1-28 <R2>
A/Cross-references: GB:558913; NID:g2929587
A/Experimental source: Sprague-Dawley
C/genetics:
A/Genes: Fst
A/Intons: 29/1; 93/1; 166/1; 241/1; 318/2/1

C:Superfamily: follistatin; kazal proteinase inhibitor homology
 F:114-164/Domain: kazal proteinase inhibitor homology <KPI1>
 F:187-239/Domain: kazal proteinase inhibitor homology <KPI2>
 F:265-316/Domain: kazal proteinase inhibitor homology <KPI3>

Query Match 16.4%; Score 113.5; DB 2; Length 344;
 Best Local Similarity 27.2%; Pred. No. 0.014;
 Matches 44; Conservative 13; Mismatches 54; Indels 51; Gaps 11;

OY 1 PNLMVDAOSTCERCSRDQCAAEKCCIN-----CI-----NVCG----- 33
 Db 154 PELEVOYGCKKTC-RDVFPGSSSTCVDDTNNAYCTCNRICPEPSSSQSLCGNDGV 212
 OY 34 -----LH-----SCVAARPPG--SPAAPTAAACGEGFVCPQOQSDCDIWD---GQPVCR-- 77
 Db 213 TYSSACHLRKATCLLGISIGLAVGKCTIKAKSCEDIQC-GGKKKC-LMDFKVGGRGCSLIC 270
 OY 78 ---CRDRCEKEPSFTCASDGLTYNNRCYMDAEACLRGLHLHI 116
 Db 271 DELCPDSKSDP--VCASDNATYASECAMKEAACSSGVLEEV 310

RESULT 12

follistatin - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
 C:Accession: S45321; S51666
 R:Albanio, R.M.; Arkell, R.; Beddington, R.S.P.; Smith, J.C.
 Development 120, 803-813, 1994
 A:Title: Expression of inhibin subunits and follistatin during postimplantation mouse de
 vel.

A:Reference number: S45321; MUID:95324368
 A:Accession: S45321
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1343 <ALB>
 A:Cross-references: EMBL:Z29533; NID:9488368; PIDN:CAA82648.1; PID:9488369
 R:Rtvoys, O.; Tuuri, T.; Erasmee, M.; Salnio, K.; Hilden, K.; Saxen, L.; Gilbert, S.
 Submitted to the EMBL Data Library, December 1994
 A:Description: Activin disrupts epithelial branching morphogenesis in developing glandu
 A:Reference number: S51666
 A:Accession: S51666
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 3-240, 'KA', 242-340 <RIT>
 A:Cross-references: EMBL:X03377; NID:9603573; PIDN:CAA58291.1; PID:9603574
 C:Superfamily: follistatin; kazal proteinase inhibitor homology
 F:114-164/Domain: kazal proteinase inhibitor homology <KPI>
 F:264-315/Domain: kazal proteinase inhibitor homology <KPI2>

Query Match 16.3%; Score 113; DB 2; Length 343;
 Best Local Similarity 26.1%; Pred. No. 0.015;
 Matches 42; Conservative 14; Mismatches 55; Indels 50; Gaps 10;

OY 1 PNLMVDAOSTCERCSRDQCAAEKCCIN-----VCG----- 33
 Db 154 PELEVOYGCKKTC-RDVFPGSSSTCVDDTNNAYCTCNRICPEPSSSQSLCGNDGV 212
 OY 34 -----LH-----SCVAARPPG--SPAAPTAAACGEGFVCPQOQSDCDIWD---GQPVCR-- 77
 Db 213 TYSSACHLRKATCLLGISIGLAVGKCTIKSCEDIQC-GGKKKC-LMDSKVGGRGCSLICD 270
 OY 78 ---CRDRCEKEPSFTCASDGLTYNNRCYMDAEACLRGLHLHI 116
 Db 271 ELCPDSKSDP--VCASDNATYASECAMKEAACSSGVLEEV 309

RESULT 13
 AGRT
 agrin - rat

C:Species: Rattus norvegicus (Norway rat)
 C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C:Accession: JH0399; A38856
 R:Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
 Neuron 6, 811-823, 1991
 A:Title: Structure and expression of a rat agrin.
 A:Reference number: JH0399; MUID:91222570
 A:Accession: JH0399

A:Molecule type: mRNA
 A:Residues: 1-1779;1799-1959 <RUP>
 A:Cross-references: GB:M64780; NID:9202798; PIDN:AAA40703.1; PID:9202800
 A:Experimental source: embryonic spinal cord
 A:Note: It is uncertain whether Met-1, Met-18, or Met-24 is the initiator
 R:Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
 J. Neurosci. 12, 3535-3544, 1992
 A:Title: Structure and chromosomal localization of the mammalian agrin gene.
 A:Reference number: A38856; MUID:92407628
 A:Accession: A38856

A:Molecule type: mRNA
 A:Residues: 1780-1798 <RU2>
 A:Cross-references: GB:S44194
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholin
 C:Comment: 90% of rat embryonic transcripts encode the variant labeled below as form
 1 (choline receptor clustering activity).

C:Superfamily: agrin; EGF homology; kazal proteinase inhibitor homology; laminin G r
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
 F:1-1959/Product: agrin, form 1 #status predicted <AG1>
 F:1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
 F:1-1779,1799-1959/Product: agrin, form 5 #status predicted <AG5>
 F:1-1779,1788-1959/Product: agrin, form 3 #status predicted <AG3>
 F:1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
 F:22-50/Region: Hydropobic

F:88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
 F:163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
 F:236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
 F:307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
 F:381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
 F:446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
 F:511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
 F:540-542/Region: motor neuron attachment (L-R-E) motif
 F:556-645/Domain: Kazal proteinase inhibitor homology <KPI8>
 F:688-739/Domain: laminin-type EGF-like homology <LE1>
 F:742-786/Domain: laminin-type EGF-like homology <LE2>
 F:814-864/Domain: kazal proteinase inhibitor homology <KPI9>
 F:869-992/Region: serine/threonine-rich
 F:1084-1086/Region: motor neuron attachment (L-R-E) motif
 F:1147-1215/Region: serine/threonine-rich
 F:1224-1257/Domain: EGF homology <EG1>
 F:1287-1442/Domain: laminin G repeat homology <LG1>
 F:1444-1476/Domain: EGF homology <EG2>
 F:1483-1515/Domain: EGF homology <EG3>
 F:1555-1706/Domain: laminin G repeat homology <LG2>
 F:1713-1747/Domain: EGF homology <EG4>
 F:1807-1959/Domain: laminin G repeat homology <LG3>
 F:97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454
 F:146,1483-1494,1488-1504,1506-1515/Dsulfide bonds: #status predicted
 F:145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.2%; Score 112; DB 1; Length 1959;
 Best Local Similarity 28.9%; Pred. No. 0.062;
 Matches 35; Conservative 14; Mismatches 48; Indels 24; Gaps 6;

OY 7 AOSTCERCS--RDQCAAEKCCINVGILHS--C-----VAARFPGSPAAPTAAASC 55
 Db 381 AECEQGVRCSTIDPVGSDGVTYGVCLESLMCTLREIOVARGPCDP-----CGCG 435
 OY 56 EGFVCPQOQSDIWDGQPVCRCDRCEKEPSFTCASDGLTYNNRCYMDAEACLRGLHLHI 115
 Db 436 -----RFGSLCEVETGR--CVPSECVESAQPVGSDGHTYASECELHVHACTHOISLY 487
 OY 116 I 116

Db 488 v 488

RESULT 14

S38251

follicle-statin-related protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 17-Nov-2000

C:Accession: S38251

R:Shibanuma, M.; Mashimo, J.; Mita, A.; Kuroki, T.; Nose, K.

Eur. J. Biochem. 217, 13-19, 1993

A:Title: Cloning from a mouse osteoblastic cell line of a set of transforming-growth-fac

A:Reference number: S38251; MUID:94039028

A:Accession: S38251

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-306 <SH1>

A:Cross-references: GB:M91380; NID:9349005; PIDN:AC37633.1; PID:9349006

C:Superfamily: Kazal proteinase inhibitor homology

F:48-96/Domain: Kazal proteinase inhibitor homology <KPI5>

Query Match

Best Local Similarity 16.1%; Score 111.5; DB 2; Length 306;

Matches 24; Conservative 16; Mismatches 30; Indels 31; Gaps 3;

QY 18 DDDCAAEKCCINV-CGLHSCVAARPPGSPAAPTAA

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

S38251

Search completed: February 26, 2002, 01:29:35
Job time: 414 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:03 : Search time 69.26 Seconds
(without alignments)
38.664 Million cell updates/sec

Title: US-09-819-136-2_COPY_39_157

Perfect score: 692
Sequence: 1 PNLWVDAQSTGERCSRDPD.....RCYMDAEACRLGLHLHYPC 119

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5b_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6a_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6b_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	16.2	1940	2	US-08-644-271-30 Sequence 30, Appl
2	104	15.0	160	3	US-09-191-647-5 Sequence 5, Appl
3	104	15.0	160	4	US-09-540-245A-5 Sequence 5, Appl
4	104	15.0	160	4	US-09-540-153-5 Sequence 5, Appl
5	101.5	14.7	263	2	US-08-972-008-2 Sequence 2, Appl
6	99	14.3	71	2	US-08-972-008-4 Sequence 4, Appl
7	99	14.3	154	3	US-09-191-647-10 Sequence 10, Appl
8	99	14.3	154	4	US-09-540-245A-10 Sequence 10, Appl
9	99	14.3	154	4	US-09-540-153-10 Sequence 10, Appl
10	97.5	14.1	2703	1	US-08-185-432-19 Sequence 19, Appl
11	96	13.9	814	4	US-09-813-819-4 Sequence 4, Appl
12	96	13.9	855	4	US-09-813-819-2 Sequence 2, Appl
13	95	13.7	374	2	US-08-820-170A-25 Sequence 25, Appl
14	95	13.7	374	3	US-09-055-699-25 Sequence 25, Appl
15	95	13.7	374	4	US-09-273-565-25 Sequence 25, Appl
16	95	13.7	374	5	PCT-US95-06385-2 Sequence 2, Appl
17	91.5	13.2	380	3	US-08-468-846-2 Sequence 2, Appl
18	91.5	13.2	380	4	US-08-915-096A-2 Sequence 2, Appl
19	91.5	13.2	4654	4	US-08-476-515A-84 Sequence 84, Appl
20	91.5	13.2	4655	4	US-08-652-877-84 Sequence 84, Appl
21	91.5	13.2	4655	4	US-08-652-877-86 Sequence 86, Appl
22	91.5	13.2	4655	4	US-08-652-877-88 Sequence 88, Appl
23	91.5	13.2	4655	4	US-08-652-877-90 Sequence 90, Appl
24	90	13.0	4544	1	US-08-469-686-52 Sequence 52, Appl
25	89.5	12.9	718	1	US-08-444-792-4 Sequence 4, Appl
26	89.5	12.9	718	1	US-08-445-042-4 Sequence 4, Appl
27	89.5	12.9	718	1	US-08-445-042-4 Sequence 4, Appl

28	89.5	12.9	788	2	US-07-728-215-32 Sequence 32, Appl
29	88	12.7	678	1	US-08-282-141-2 Sequence 2, Appl
30	88	12.7	678	1	US-08-435-434-2 Sequence 2, Appl
31	88	12.7	678	1	US-08-435-436-2 Sequence 2, Appl
32	88	12.7	678	2	US-08-438-863-2 Sequence 2, Appl
33	88	12.7	678	2	US-08-438-864-2 Sequence 2, Appl
34	88	12.7	678	3	US-08-438-862-2 Sequence 2, Appl
35	88	12.7	678	4	US-08-628-747-2 Sequence 2, Appl
36	88	12.7	678	4	US-08-402-253-2 Sequence 2, Appl
37	88	12.7	678	4	US-08-443-866B-2 Sequence 2, Appl
38	87	12.6	2523	1	US-08-185-432-18 Sequence 18, Appl
39	86.5	12.5	1111	1	US-08-317-450B-15 Sequence 15, Appl
40	86.5	12.5	1111	4	US-08-800-593-15 Sequence 15, Appl
41	86.5	12.5	1193	1	US-08-317-450B-13 Sequence 13, Appl
42	86.5	12.5	1193	4	US-08-800-593-13 Sequence 13, Appl
43	86	12.4	798	2	US-07-728-215-30 Sequence 30, Appl
44	85.5	12.4	680	1	US-08-211-430-2 Sequence 2, Appl
45	85.5	12.4	680	3	US-08-761-136-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644, 271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008, 657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobeart, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SRO ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rat Agt1n
; LOCATION: 1..1940
; OTHER INFORMATION:
; US-08-644-271-30

Db 134 DNGILLYN 141

RESULT 5

US-08-972-008-2
Sequence 2, Application US/08972008
Patent No. 5942420

GENERAL INFORMATION:

APPLICANT: Holtzman, Douglas A.
TITLE OF INVENTION: No. 5942420el Molecules of the Follistatin-Related

Protein

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street

CITY: Boston
STATE: Massachusetts

COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/972,008
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: MNI-026
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 263 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-972-008-2

Query Match

Best Local Similarity 14.7%; Score 101.5; DB 2; Length 263;
Matches 36; Conservative 10; Mismatches 46; Indels 41; Gaps 8;

Db 4 WVDQSTCERCS--RDQCAAEKCC-----INVGLHSCVAA 40

Db 39 WL--QGOGEATCSLVLTQDTYRAE--CCASGNIDTAMSLTHPGKNINLGLVHCLPC 95

Db 41 RFGSPAPPTTAAECGFCVPCQSGDDIDWDGQPVRCRDRCEKPS--FTCASDGLTYV 98

Db 96 K-----DSCDGEVC-GPGKACRMGLGRPRCECAPDCSGLPARLYVCGSDGATYR 143

Db 99 NRCYDAEACLRG 111

Db 144 DECELRARC-RG 155

RESULT 6

US-08-972-008-4
Sequence 4, Application US/08972008
Patent No. 5942420

GENERAL INFORMATION:

APPLICANT: Holtzman, Douglas A.
TITLE OF INVENTION: No. 5942420el Molecules of the Follistatin-Related

Protein

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street

CITY: Boston

STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/972,008
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: MNI-026
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 71 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

US-08-972-008-4

Query Match

Best Local Similarity 14.3%; Score 99; DB 2; Length 71;
Matches 23; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

Db 54 SCGEFVCPQSGDDIDWDGQPVRCRDRCEKPS--FTCASDGLTYNRCYDAEACLRG 111

Db 2 SCDGEVC-GPGKACRMGLGRPRCECAPDCSGLPARLYVCGSDGATYRDECELRARC-RG 59

RESULT 7

US-09-191-647-10
Sequence 10, Application US/09191647
Patent No. 6046015

GENERAL INFORMATION:

APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas

APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions

FILE REFERENCE: 898-031-3
CURRENT APPLICATION NUMBER: US/09/191,647

CURRENT FILING DATE: 1998-11-13
EARLIER APPLICATION NUMBER: 60/065,544

EARLIER FILING DATE: 1997-11-14
EARLIER APPLICATION NUMBER: 60/081,057

EARLIER FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 10

TYPE: PRT
ORGANISM: mouse

US-09-191-647-10

Query Match

Best Local Similarity 14.3%; Score 99; DB 3; Length 154;
Matches 27; Conservative 12; Mismatches 28; Indels 18; Gaps 6;

Db 15 CSRDQCAAEKCC-----CINVCGLHSCVAA-RFGS-----PAAFTTAAECGFCVPCQSG 64

Db 20 CSENDQDCRDKRCQNGACVDEVNSYACLCYEGYSGOLCEIIPAP--RSSCGTEC-QNG 76

Db 65 SDCDIDWDGQPVRCR-----RDRCEK 84

Db 77 ANCVDSGRVCCQLPFGGPECEK 101

RESULT 8

US-09-540-245A-10
; Sequence 10, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-245A-10

Query Match 14.3%; Score 99; DB 4; Length 154;
Best Local Similarity 31.8%; Pred. No. 0.017;
Matches 27; Conservative 12; Mismatches 28; Indels 18; Gaps 6;

Qy 15 CSRDODCAAEK-----CINVCGLHSCVAAR-FPGS-----PAPPTAASCEGFVCPQOG 64
Db 20 CSENDDCKDKHCQNGACQVDEYNSTACLCVEGYSGLCEIPIAP--RSSCEGTEC-QNG 76
Qy 65 SDCDIMDGPVRC-----RDRCEK 84
Db 77 ANCVDSGRVCCQLPFGGPECEK 101

RESULT 9

US-09-540-153-10
; Sequence 10, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-153-10

Query Match 14.3%; Score 99; DB 4; Length 154;
Best Local Similarity 31.8%; Pred. No. 0.017;
Matches 27; Conservative 12; Mismatches 28; Indels 18; Gaps 6;

Qy 15 CSRDODCAAEK-----CINVCGLHSCVAAR-FPGS-----PAPPTAASCEGFVCPQOG 64
Db 20 CSENDDCKDKHCQNGACQVDEYNSTACLCVEGYSGLCEIPIAP--RSSCEGTEC-QNG 76
Qy 65 SDCDIMDGPVRC-----RDRCEK 84
Db 77 ANCVDSGRVCCQLPFGGPECEK 101

RESULT 10

US-08-185-432-19
; Sequence 19, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Attavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle J.
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2703 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-19

Query Match 14.1%; Score 97.5; DB 1; Length 2703;
Best Local Similarity 21.8%; Pred. No. 0.46;
Matches 42; Conservative 15; Mismatches 51; Indels 85; Gaps 11;

Qy 4 WYDAOSTCE-----RECSRDOD-CAAEK-----CINVCGLHSCVAAR-FPGSPAAPT 51
Db 888 FLDFSCCTCKLGYTRYCEDIDECSLSPCRNGASCLWVPSYRCLCTKGYEGRCALNT 947
Qy 52 AASCEGFVCPQOGSDCDIMDGPVRCRD-----RCEKEP-----S 87
Db 948 -DDCASFP-C-ONGRTCLDIDISDYLSCVDFGDKHCETDINECLSQPONGATCSQYVNS 1005
Qy 88 FTGAS-----DELTVY-----NRCY 102
Db 1006 YTCYPLGFSGINCOTNDECTESSCLNGSCIDISINGNCISLAGYSGANCQYKLNK- 1064
Qy 103 MDAEACLRGLRH 115

	Query Match	13.9%	Score 96;	DB 4;	Length 855;
	Best Local Similarity	34.0%	Pred. No. 0.19;	Mismatches 37;	Indels 18; Gaps 7;
	Matches	33;	Conservative	9;	
Qy	10 TCERCSRD-----ODCA--AAEKCCINVGSLGSVAARFPESPAPPTAASCEG-FV	59			
	: : :				
Db	392 TCIIEASTDPLPGINFSCNRRALEKLLD--GGSGCLEERL--PSLPMAAFCGMFEV	446			
Qy	60 CPOGSDCDIWD--GPVCRCRDRCKEKEPFTCASDG	94			
	: :				
Db	447 EPGQCDCGFLUDCVDPCCSL-TCQLRPGACQCADSG	482			

RESULT 14
US-09-055-699-25
; Sequence 25, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sugihara, Mion, Zinn, Macpeak & Sears
; STREET: 2100 Pennsylvania Avenue, N.W.

```

1      CITY: Washington
2      STATE: D.C.
3      COUNTRY: United States
4      ZIP: 20037-3202
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: IBM PC compatible
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: PatentIn Release #1.0,
11     CURRENT APPLICATION DATA:
12     APPLICATION NUMBER: US/09/055,699
13     FILING DATE:
14     CLASSIFICATION:
15
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER: 08/820,170
18     FILING DATE:
19
20     TELECOMMUNICATION INFORMATION:
21     TELEPHONE: (202) 293-7060
22     TELEFAX: (202) 293-7860
23     TELEX: 6491103
24
25     INFORMATION FOR SEQ ID NO: 25:
26     SEQUENCE CHARACTERISTICS:
27     LENGTH: 374 amino acids
28     TYPE: amino acid
29     TOPOLOGY: linear
30
31     MOLECULE TYPE: protein
32
33     US-09-055-699-25

```

Query Match	13.7%;	Score 95;	DB 3;	Length 374;
Best Local Similarity	25.2%;	Pred. No. 0.1;		
Matches 32;	Conservative 20;	Mismatches 49;	Indels 26;	Gaps 6

QY	67	-----CDI-----	-PVCRCRGRCEKEEFTCASDGLTYNNCYMDAENCLR	110
Dd	64	NDFLECDTNNCKEDGDECLRIIGDTVTCVCFRCNNDYVPVCGSNESGYONECYLROAAKQ		123
QY	111	GLHLHIY	117	
Dd	124	OSETILVV	130	

RESULT 15
US-09-273-565-25
; Sequence 25, Application US/09273565A
; Patent No. 6166100

```

?
? GENERAL INFORMATION:
? APPLICANT: FUJIWARA, TSUTOMU
? APPLICANT: WATANABE, TAKESHI
? APPLICANT: HORIE, MASATO
? TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
? TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONUGATING ENZYME
? FILE REFERENCE: Q-53559
? CURRENT APPLICATION NUMBER: US/09/273,565A
? CURRENT FILING DATE: 1999-03-22
? EARLIER APPLICATION NUMBER: 09/055,699
? EARLIER FILING DATE: 1998-04-07
? EARLIER APPLICATION NUMBER: 08/820,170
? EARLIER FILING DATE: 1997-03-19
? EARLIER APPLICATION NUMBER: JP 63410/1996
? EARLIER FILING DATE: 1996-03-19
? EARLIER APPLICATION NUMBER: JP 69163/1997
? EARLIER FILING DATE: 1997-03-05
? NUMBER OF SEQ ID NOS: 95
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 25
? LENGTH: 374
?
? TYPE: prt
?
? ORGANISM: Homo sapiens

```

US-09-273-565-25

Query Match	13.7%	Score 95;	DB 4;	Length 374;
Best Local Similarity	25.2%	Pred. No. 0.1;		
Matches 32; Conservative	20;	Mismatches 49;	Indels 26;	Gaps 6;

QY	13	RECRDDDCAAAEKCCINVCGLHSCVAPRPGSPAATTAASE---	GFPQPOGSP---	66
Dd	8	RCCSMTLCEGF--CWLLLPVMLLIARBPVKLAAPTSLSDQGTPGMWC-	-SGYDRE	63
QY	67	-----CII-----WDG-----PVCRCRCRCEKEPSFTCASGLTYNRCYMADAELCLR		110
Dd	64	NDFLELCPTNCKPKPEGLRLIGDYIVTCVQCFKCNRNDIYPVCGSGESTIQNECTLRQAACKO		123
QY	111	GLHLHIY	117	
Dd	124	QSELIIV	130	

Search completed: February 26, 2002, 01:28:04
Job time: 358 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:14 ; Search time 144.8 Seconds
(without alignments)
65,661 Million cell updates/sec

Title: US-09-819-136-2_COPY_93_157
Perfect score: 390
Sequence: 1 CEGFVCPQGSDDCDINDGPR.....RCYMDAACLRLGHLHTVPC 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	120	30.8	1296	5 Q22452	Q22452 caenorhabditis
2	114	29.2	2026	4 Q00468	Q00468 homo sapien
3	113	29.0	315	13 Q9W600	Q9W600 gallus galli
4	110	28.2	299	13 Q9DGB0	Q9DGB0 xenopus lae
5	110	28.2	306	11 Q99J19	Q99J19 mus musculu
6	110	28.2	308	6 Q9GKY0	Q9GKY0 macaca fasc
7	103	26.4	654	5 Q9VSI9	Q9VSI9 drosophila
8	98	25.1	256	11 Q99PW7	Q99PW7 rattus norv
9	96	24.6	256	11 Q9EQC7	Q9EQC7 mus musculu
10	95	24.4	263	4 Q95633	Q95633 homo sapien
11	93	23.8	319	13 Q91376	Q91376 xenopus lae
12	91	23.3	773	4 Q9NSW7	Q9NSW7 homo sapien
13	91	23.3	850	4 Q9ULF7	Q9ULF7 homo sapien
14	88	22.6	343	13 Q90844	Q90844 gallus galli
15	88	22.6	343	13 Q9PS97	Q9PS97 gallus galli
16	87	22.3	344	4 Q9BTH0	Q9BTH0 homo sapien
17	84	21.5	354	11 Q9JUS1	Q9JUS1 mus musculu
18	84	21.5	373	11 Q9OYV1	Q9OYV1 rattus norv
19	84	21.5	380	4 Q13086	Q13086 homo sapien

ALIGNMENTS

RESULT	ID	Q22452	PRELIMINARY	PRT	1296 AA.	
AC	Q22452					
DT	01-NOV-1996	(TREMBLrel. 01, Created)				
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)				
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)				
DE	SIMILAR TO AGRIN AND FOLLISTATIN.					
GN	T13C2.5.					
OS	Caenorhabditis elegans.					
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;					
OC	Rhabditidae; Pelodierinae; Caenorhabditis.					
OX	NCBI_TaxID=6239;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=94150718; PubMed=7906398;					
RA	Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,					
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,					
RA	Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fulton L.,					
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,					
RA	Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,					
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,					
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,					
RA	Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,					
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,					
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,					
RT	2.2 Mb of contiguous nucleotide sequence from chromosome III of C.					
RT	elegans."					
RL	Nature 368:32-38(1994).					
RP	[2]					
RP	SEQUENCE FROM N.A.					
RA	Du Z.;					
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RA	Waterston R.;					
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.					
DR	EMBL: U40030; AA81133.1; -.					
DR	HSSP: P37109; IPCE.					
DR	InterPro: IPR000561; EGF-like.					
DR	InterPro: IPR003645; F0LN.					
DR	InterPro: IPR002350; kazai.					

DR InterPro: IPR001239; Kazal_inhib.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam: PF000050; Kazal; 9.
 DR Pfam: PF000053; Laminin_EGF; 2.
 DR PRINTS: PR00290; KAZALINHTR.
 DR SMART: SM00180; EGF_Lam; 1.
 DR SMART: SM00001; EGF_Like; 1.
 DR SMART: SM00274; FOLN; 4.
 DR SMART: SM00280; KAZAL; 9.
 DR SMART: SM00282; Lamg; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
 DR GLYCOPROTEIN: Laminin EGF-like domain; Repeat.
 SO SEQUENCE 1296 AA; 145178 MW; 05094BC185839690 CRC64;

Query Match 30.8%; Score 120; DB 5; Length 1296;
 Best Local Similarity 36.4%; Pred. No. 1.3e-06;
 Matches 20; Conservative 9; Mismatches 24; Indels 2; Gaps 2;

OY 1 CEGFVCPQGSDDCIIM-DGQVPCRCRDRCERPEFTCASDGLITYNRCYMDAEC 54
 Db 295 CHGKCP-NGQTQDGYDRPECKSECTMNSAHVCTGKTYLNECFKLIAAC 348

RESULT 2
 ID 000468 PRELIMINARY; PRT; 2026 AA.
 AC 000468;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE AGRIN PRECURSOR (FRAGMENT).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
 RA Verkamp J.H., Monnens L.A.H., van den Heuvel L.P.W.J.;
 RL Eur. J. Biochem. 0:0-0(1998).
 RN [2]
 RP SEQUENCE OF 1-153 FROM N.A.
 RX MEDLINE-96224170; PubMed-8617505;
 RA Lennon G., Aufray C., Polymeropoulos M., Soares M.B.;
 RT "The I.M.A.G.E. Consortium: an integrated molecular analysis of
 RT genomes and their expression.";
 RL Genomics 33:151-152(1996).
 RN [3]
 RP SEQUENCE OF 1-153 FROM N.A.
 RA Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
 RL J. Cell Biol. 0:0-0(0).
 DR EMBL: AF016903; AAC39776.1; -
 DR EMBL: U84406; AAB52917.1; -
 DR HSSP: P37109; IPCE.
 DR InterPro: IPR000561; EGF_1like.
 DR InterPro: IPR001438; EGF_1I.
 DR InterPro: IPR003884; FACLMAC.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR002350; Kazal.
 DR InterPro: IPR001239; Kazal_inhib.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR000082; SEA.
 DR InterPro: IPR001455; UPF0033.
 DR Pfam: PR00008; EGF; 4.
 DR Pfam: PF00050; Kazal; 9.
 DR Pfam: PF00053; Laminin_EGF; 2.
 DR Pfam: PF00054; Laminin_G; 3.

DR Pfam: PF01390; SEA; 1.
 DR PRINTS: PR00010; EGFBL00D.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PRINTS: PR00290; KAZALINHTR.
 DR SMART: SM00180; EGF_Lam; 2.
 DR SMART: SM00001; EGF_Like; 4.
 DR SMART: SM00057; FIMAC; 4.
 DR SMART: SM00274; FOLN; 5.
 DR SMART: SM00280; KAZAL; 9.
 DR SMART: SM00282; Lamg; 3.
 DR SMART: SM00200; SEA; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_6.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE: PS01148; UPF0033; UNKNOWN_1.
 KW EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat;
 KW Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 10 POTENTIAL.
 FT CHAIN 11 2026 AGRIN.
 SO SEQUENCE 2026 AA; 212881 MW; 4AB0E710CD4B8EF CRC64;

Query Match 29.2%; Score 114; DB 4; Length 2026;
 Best Local Similarity 33.8%; Pred. No. 1.1e-05;
 Matches 23; Conservative 9; Mismatches 32; Indels 4; Gaps 2;

OY 1 CEGFVCPQGSDDCIIMDGQVPCRCRDRCERPEFTCASDGLITYNRCYMDAEC 60
 Db 449 CLGVQC-AFGATCAVKNQGAACECLQACSSLYDPVCGSDVTVGACELTACTLGREI 507

OY 61 HIV--PC 65
 Db 508 QVARKGPC 515

RESULT 3
 ID Q9W600 PRELIMINARY; PRT; 315 AA.
 AC Q9W600;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE FLIK PROTEIN.
 DE FLIK.
 GN FLIK.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Patel K.;
 RT "Cloning and early dorsal-axial expression of Flik, a chick
 RT follistatin-related gene: evidence for involvement in
 RT dorsalisation/neural induction.";
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96427388; PubMed-8812133;
 RA Patel K., Connolly D., Anthor H., Nose K.;
 RT "Cloning and early dorsal axial expression of Flik, a chick
 RT follistatin-related gene: evidence for involvement in
 RT dorsalisation/neural induction.";
 RL Dev. Biol. 178:327-342(1996).
 DR EMBL: AJ238977; CAB42968.1; -
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR002350; Kazal.
 DR InterPro: IPR003645; FOLN.
 DR Pfam: PF00050; Kazal; 1.
 DR SMART: SM00274; FOLN; 1.
 DR SMART: SM00280; KAZAL; 1.
 SO SEQUENCE 315 AA; 35815 MW; C01889E005658A67 CRC64;

Query Match	29.08;	Score 113;	DB 13;	Length 315;
Best Local Similarity	31.7%;	Pred. NO. 2.6e-06;		
Matches 20;	Conservative 14;	Mismatches 27;	Indels 2;	Gaps 2;

Oy	1	C E F E V C E O O G S D P D I W D - G Q P V C R C R D R C K E P S F T C A S G L Y Y N N C Y A D A E A C L R G L H	59
		: : : : : : : : : : : : : : : : : : : :	
Dd	31	C A N V F C - G R G A E C A V T E K E P F C L C I E D C K R H G R P V C G S N G K T Y L N H C E L H R D A C L T G S K	89
Oy	60	L H I	62
Dd	90	I Q V	92

RESULT	4
Q9DGB0	
ID	Q9DGB0
AC	Q9DGB0;
DT	01-MAR-2001 (TREMBLrel, 16, Created)
DT	01-MAR-2001 (TREMBLrel, 16, Last sequence update)
DT	01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE	FOLLISTATIN-RELATED PROTEIN.
GN	XFRP.
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OX	Xenopodinae; Xenopus.
OX	NCBI_taxid=8355;
RN	[1]
RA	SEQUENCE FROM N.A.
RA	Okabayashi K., Shoji H., Asashima M., Sugino H.;
RL	Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99121187; PubMed=9920730;
RA	Okabayashi K., Shoji H., Onuma Y., Nakamura T., Nose K., Sugino H.,
RA	Asashima M.;
RT	"cDNA Cloning and Distribution of the Xenopus Follistatin-Related
RT	Protein.";
RL	Biochem. Biophys. Res. Commun. 254:42-48(1999).
CC	- SIMILARITY: TO EF-HAND FAMILY.
DR	EMBL; AB049354; BAB13800.1; -.
DR	InterPro; IPR002048; EF-hand.
DR	InterPro; IPR003645; FOLN.
DR	InterPro; IPR002350; Kazal.
DR	Pfam; PF000036; efhand; 2.
DR	Pfam; PF000050; kazal; 1.
DR	SMART; SMO0274; FOLN; 1.
DR	SMART; SMO0280; KAZAL; 1.
DR	Calcium-binding.
QO	SEQUENCE 299 AA; 33701 MW; 43ADEF06F5E62AF CRC64;

[illegible]

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FOLLISTATIN-LIKE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC006185, AAH06185.1; -
SO SEQUENCE 306 AA; 34534 MW; 0AEB11C321EB8FDC CRC64;

Query Match	28.2%	Score 110;	DB 11;	Length 306;
Best Local Similarity	31.7%	Pred. No. 5.9e-06;		
Matches	20;	Conservative 12;	Mismatches 29;	Indels 2;
				Gaps 2;

Oy 1 CEGPVCQQQSGDPCDIWD -GAPVCRDRRCCEKPSCTCASDGLTFYNNCYADAEACLRGLH 59
 : : : : : : : : : : : : : : : : : :
Db 29 CANVFC-GAGRECAVTEKEPFCTCBQCPRFKRKYVCGSSNOKTYLNHCHELROACLTSKR 87

Oy 60 LHI 62
 : :
Db 88 IQV 90

RESULT	6			
09GKY0				
ID	09GKY0	PRELIMINARY;	PRT;	308 AA.
AC	09GKY0;			
DT	01-MAR-2001	(TrEMBLrel. 16, Created)		
DT	01-MAR-2001	(TrEMBLrel. 16, last sequence update)		
DT	01-JUN-2001	(TrEMBLrel. 17, last annotation update)		
DE	OCCL.			
GN	OCCL			
OS	Macaque fascicularis (Crab eating macaque) (Cynomolgus monkey).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheraia; Primates; Catarrhini; Cercopithecoidea;			
OC	Cercopithecinae; Macaca.			
OX	NCBI_TaxID=9541;			
OX	[1]			
RN				
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RA	Tochtetani S.;			
RT	"Neuronal heterogeneity in macaque neocortex revealed by occl mRNA			
RT	activity-dependent expression."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: TO EF-HAND FAMILY.			
DR	EMBL; AB039661; BAB20770.1; ..			
DR	InterPro: IPR002048; EF-hand.			
DR	InterPro: IPR003645; FOLN.			
DR	InterPro: IPR002350; kazal.			
DR	InterPro: IPR001007; VMFC.			
DR	Pfam: PF00036; ehand; 2.			
DR	Pfam: PF00050; kazal; 1.			
DR	SMART: SM00274; FOLN; 1.			
DR	SMART: SM00280; KAZAL; 1.			
DR	SMART: SM00214; VMC; 1.			
DR	SMART: SM00011; VMC_def; 1.			
DR	Calcium-binding.			
Q0	SEQUENCE 308 AA; 34999 MW; 4B283609CDEF535D9 CRC64;			

```

Query Match      28.2%; Score 110; DB 6; length 308;
Best Local Similarity 31.7%; Pred. No. 6e-06;
Matches 20; Conservative 12; Mismatches 29; Indels 2; Gaps 2;
OY   1 CEGFYCPOGSDCDIID-GQPYCRCDRCCEKPEPSTCA SDGLTYNNRCTMDAEACLRGIH 59
          : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    31 CANVFC-GAGRCAYTEGEPTCLCI EOCKPKHKRPVCGSGNGKTYLNHC ELHRADACTGSK 89
OY   60 LHI 62

```

Db 90 10V 92

RESULT 7

09VSJ9 PRELIMINARY; PRT: 654 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CG7159 PROTEIN.

GN CG7159.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,

RA Foster K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.W., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003555; AAF50420.1; -

DR HSSP; P01003; 10VO.

DR FlyBase; FBgn0035893; CG7159.

DR InterPro; IPR002350; kazal.

DR Pfam; PF00050; kazal; 9.

DR SMART; SM00280; KAZAL; 8.

SO SQUENCE 654 AA; 71626 MW; 515582E86FC5048C CRC64;

Query Match 26.4%; Score 103; DB 5; Length 654;

Best Local Similarity 30.0%; Pred. No. 8.9e-05;

Matches 24; Conservative 9; Mismatches 25; Indels 22; Gaps 4;

Qy 5 VCPQGDSC-----DIMDGPVCR-----CRDCEKEPFTCASDGLTYNRCYM 49

Db 182 ICGLRKRTCSRGVSLINDVRG---CERSKSDCKHRCSTKEDPVCGTGDRTYLRML 238

Qy 50 DAACLRLH---HIVPC 65

Db 239 RVQSCRVGLAAVKLSHVGPC 258

RESULT 8

099PW7 PRELIMINARY; PRT: 256 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE FOLLISTATIN-RELATED PROTEIN FLRG PRECURSOR.

GN FLRG.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RA Uchiyama Y., Ohsawa Y., Kametaka S.;

RT "rat follistatin-related protein FLRG (rat PCF35, PC12 cell derived

RT tropic factor 35).";

RL Submitted (DRC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB021295; BAB32664.1; -

KM Signal.

FT SIGNAL

SO SEQUENCE 256 AA; 27109 MW; C8464166C0A1D56F CRC64;

Qy 1 CEGFVCPQGDSCDIMDGPVCRDCEKEPFTCASDGLTYNRCYMDAECARG 57

Db 97 CDGVECGPGKACMGAGRHCEVSNCEGVAFQVCGSDGATYRDECELTARC-RG 153

RESULT 9

ID 09EOC7 PRELIMINARY; PRT: 256 AA.

AC 09EOC7;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE FOLLISTATIN-LIKE PROTEIN (FOLLISTATIN-RELATED PROTEIN FLRG PRECURSOR).

GN FLRG.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6;

RX PubMed=11010968;

RA Tsuchida K., Arai K.Y., Kuramoto Y., Yamakawa N., Hasegawa Y.,

RA Sugino H.;

RT "Identification and characterization of a novel Follistatin-like

RT protein as a binding protein for the TGF-beta Family.";

RL J. Biol. Chem. 275:40788-40796(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Uchiyama Y., Ohsawa Y., Kametaka S.;

RT "mouse follistatin-related protein FLRG (mouse PCF35, PC12 cell

RT derived tropic factor 35).";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Tortorello D.V., Pan Y., Sidis Y., Holtzman D., Holmes W.,
RA Keutman H., Schneyer A.;
RT "Human follistatin-related protein (FSRP) is an activin-binding homolog
of follistatin that localizes to the nucleus.";
RL Submitted (JCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF276238; AAC47666.1; -
DR EMBL: AB024429; BAB32663.1; -
DR EMBL: AF310616; AAC53666.1; -
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00050; kazal; 2.
DR SMART: SM00274; FOLN; 2.
DR SMART: SM00280; KAZAL; 2.
KM Signal.
FT SIGNAL.
SQ SEQUENCE 256 AA; 27271 MW; 58544A8BC1C02EB9 CRC64;

Query Match 24.6%; Score 96; DB 11; Length 256;
Best Local Similarity 37.3%; Pred. No. 0.00028;
Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

QY 1 CEEGVCPQGGSDCDIMDGPVCRDRCEKEPS--FTCASDGLTYNRCYMDAACLRG 57
Db 99 CDGVEG-GPGKACRMILGGRPCCECAPDCSGLPARLVGCSGSGATYRDECELRAR-C- 155

RESULT 10
QY95633 PRELIMINARY; PRT; 263 AA.
AC 095633;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FOLLISTATIN-RELATED PROTEIN FLRG (FOLLISTATIN-Like 3) (SECRETED
DE GLYCOPROTEIN).
GN FLRG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=98334491; PubMed=9671416;
RA Hayette S., Gadoux M., Martel S., Bertrand S., Tligaud I., Magaud J.P.,
RA Rimokh R.;
RT "FLRG (follistatin-related gene), a new target of chromosomal
RT rearrangement in malignant blood disorders.";
RL Oncogene 16:2949-2954(1998).
RN 12;
RP SEQUENCE FROM N.A.
RC TISSUE=CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U76702; AAC64321.1; -
DR EMBL: BC005839; AAH05839.1; -
DR HSSP: P37109; IPCE.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00050; kazal; 2.
DR SMART: SM00274; FOLN; 2.
DR SMART: SM00280; KAZAL; 2.
SQ SEQUENCE 263 AA; 27663 MW; 6A9AB86ADDAFD09C CRC64;

Query Match 24.4%; Score 95; DB 4; Length 263;
Best Local Similarity 37.3%; Pred. No. 0.00038;
Matches 22; Conservative 6; Mismatches 27; Indels 4; Gaps 3;

QY 1 CEEGVCPQGGSDCDIMDGPVCRDRCEKEPS--FTCASDGLTYNRCYMDAACLRG 57
Db 99 CDGVEG-GPGKACRMILGGRPCCECAPDCSGLPARLVGCSGSGATYRDECELRAR-C- 155

RESULT 11
QY91376 PRELIMINARY; PRT; 319 AA.
AC 091376;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FOLLISTATIN.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN 11;
RP SEQUENCE FROM N.A.
RX MEDLINE=94221645; PubMed=8168135;
RA Hemmati-Brivanlou A., Kelly O.G., Melton D.A.;
RT "Follistatin, an antagonist of activin, is expressed in the Spemann
organizer and displays direct neutralizing activity.";
RL Cell 77:283-293(1994).
DR EMBL: S69801; AAB30638.1; -
DR HSSP: P09486; IBMO.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00050; kazal; 3.
DR SMART: SM00274; FOLN; 3.
DR SMART: SM00280; KAZAL; 3.
SQ SEQUENCE 319 AA; 34914 MW; D2CF5E72E30E8EA9 CRC64;

Query Match 23.8%; Score 93; DB 13; Length 319;
Best Local Similarity 36.8%; Pred. No. 0.0008;
Matches 25; Conservative 5; Mismatches 30; Indels 8; Gaps 5;

QY 1 CEEGVCPQGGSDCDIMD--GQVCR-CRDR--EKEPSFTCASDGLTYNRCYMDAEC 54
Db 245 CEDTQC-SAGKKC-LMDSRVGRGKALCDDLCGSKSDTYTCASDNTTYPECCAMKQAC 302

QY 55 LRGHLHI 62
Db 303 STGILLEV 310

RESULT 12
QY9NSW7 PRELIMINARY; PRT; 773 AA.
AC 09NSW7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 87.2 KDA PROTEIN (FRAGMENT).
GN DKFZP566D234.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11;
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC -1- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL: AL137695; CAB70877.1; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002350; kazal.
DR Pfam: PF00036; ehand; 1.

RESULT 14

Query Match	22.6%;	Score 88;	DB 13;	Length 343;
Best Local Similarity	35.3%;	Pred. No.	0.0036;	

Matches 24; Conservative 6; Mismatches 30; Indels 8; Gaps 5;
QY 1 CEGFVCPQOQSDCDIWD---GQPVCR-CRDRG--EKPSFTCASDGLTYYNRCYMDAEAC 54
DB 244 CEDIOC-SAGKRC-LMDFKVGRCALCDLCPESKSDAEAVCASDNTTTPSECAMKEAAC 301
QY 55 LRGHLHLHI 62
DB 302 SMGVLLLEV 309

Search completed: February 26, 2002, 01:49:16
Job time: 1265 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:18 ; Search time 144.8 seconds
(without alignments)
84.854 Million cell updates/sec

Title: US-09-819-136-2_COPY_203_286
Perfect score: 442
Sequence: 1 ASLHCVSGRPPAVTWKQ.....NAAGLRADPPLSVQREPA 84

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_TODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	28.7	1248	6 Q9XT41	Q9XT41 cercopithec
2	126	28.5	814	13 Q91897	Q91897 xenopus lae
3	125.5	28.4	497	4 Q9BXN7	Q9BXN7 homo sapien
4	125.5	28.4	504	4 Q9HAD7	Q9HAD7 homo sapien
5	123	27.8	538	11 Q9QYQ7	Q9QYQ7 mus musculu
6	123	27.8	1259	11 Q9QY38	Q9QY38 mus musculu
7	121.5	27.5	379	11 Q9CWM1	Q9CWM1 mus musculu
8	121.5	27.5	492	11 Q9EM54	Q9EM54 mus musculu
9	121.5	27.5	2828	4 Q9NR99	Q9NR99 homo sapien
10	120.5	27.3	285	4 Q43608	Q43608 homo sapien
11	120.5	27.3	1380	4 Q9HCK4	Q9HCK4 homo sapien
12	120	27.1	1060	11 Q9QZ13	Q9QZ13 rattus norv
13	119	26.9	1344	11 Q9Z214	Q9Z214 mus musculu
14	117.5	26.6	1252	11 Q9JLT1	Q9JLT1 mus musculu
15	117.5	26.6	1252	11 Q9EQS9	Q9EQS9 mus musculu
16	117.5	26.6	1253	11 Q9EQS8	Q9EQS8 mus musculu
17	117	26.5	483	4 Q9UF14	Q9UF14 homo sapien
18	117	26.5	1094	4 Q9BY88	Q9BY88 homo sapien
19	116.5	26.4	1612	11 Q89026	Q89026 mus musculu

20	116.5	26.4	1651	4 Q9Y6N7	Q9Y6N7 homo sapien
21	116	26.2	1419	13 Q98SW3	Q98SW3 brachydanto
22	115.5	26.1	1651	11 Q55005	Q55005 rattus norv
23	115	26.0	433	5 Q9V644	Q9V644 drosophila
24	115	26.0	1065	4 Q94898	Q94898 homo sapien
25	115	26.0	2783	5 P91255	P91255 caenorhabd
26	114.5	25.9	1395	5 Q44924	Q44924 drosophila
27	114.5	25.9	2164	13 Q9IAR9	Q9IAR9 gallus gall
28	114.5	25.9	4162	13 Q98918	Q98918 gallus gall
29	114.5	25.9	5198	5 Q76518	Q76518 caenorhabd
30	114.5	25.9	26526	4 Q10466	Q10466 homo sapien
31	114	25.8	796	13 Q91287	Q91287 pleurodeles
32	114	25.8	1033	5 Q24327	Q24327 drosophila
33	114	25.8	1033	5 Q9V643	Q9V643 drosophila
34	114	25.8	1822	4 Q9ULT7	Q9ULT7 homo sapien
35	113	25.6	740	11 Q9D332	Q9D332 mus musculu
36	113	25.6	868	11 Q62838	Q62838 rattus norv
37	113	25.6	869	4 Q15146	Q15146 homo sapien
38	113	25.6	1021	5 Q9V430	Q9V430 drosophila
39	112.5	25.5	145	4 Q9UGA0	Q9UGA0 homo sapien
40	112.5	25.5	385	4 Q9UQF5	Q9UQF5 homo sapien
41	112.5	25.5	404	4 Q9Y3E9	Q9Y3E9 homo sapien
42	112.5	25.5	772	4 Q9Y2J6	Q9Y2J6 homo sapien
43	112.5	25.5	2000	6 Q97791	Q97791 oryctolagus
44	112.5	25.5	2221	5 Q9U1M1	Q9U1M1 drosophila
45	111.5	25.2	4370	4 Q9H3V5	Q9H3V5 homo sapien

ALIGNMENTS

RESULT	ID	Q9XT41	PRELIMINARY:	PRT:	1248 AA.
AC	Q9XT41				
AD	Q9XT41				
DT	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	NEURAL CELL ADHESION MOLECULE L1.				
GN	L1CAM.				
OS	Cercopithecus aethiops (Green monkey) (Grivet).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;				
OC	Cercopithecoidea; Cercopithecus.				
OX	NCBI_TaxID=9534;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=KIDNEY;				
RA	Veske A., Michelson P., Finckh U., Gal A.;				
RT	"Characterization of the new L1CAM mutations in patients with CRASH				
RT	syndrome."				
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX				
CC	DOMAIN				
DR	EMBL: AF129167; AAD28610.1; -.				
DR	HSPF: P20241; ICFB.				
DR	InterPro: IPR003961; FN_III.				
DR	InterPro: IPR003598; IG_C2.				
DR	InterPro: IPR003600; IG_1like.				
DR	InterPro: IPR003006; IG_MHC.				
DR	Pfam: PF00041; fn3; 5.				
DR	Pfam: PF00047; ig; 6.				
DR	SMART: SM00060; FN3; 4.				
DR	SMART: SM00408; IGc2; 5.				
DR	SMART: SM00410; IG_1like; 1.				
SO	SEQUENCE 1248 AA; 138901 MW; 2438A93A29C3BD61 CRC64;				

Query Match 28.7%; Score 127; DB 6; Length 1248;
Best Local Similarity 42.5%; Pred. No. 2.3e-06;
Matches 34; Conservative 5; Mismatches 35; Indels 6; Gaps 2;
QY 1 ASLHCVSGRPPAVTWKQSHQRENLIMRPPQMGVNVVTSIGQLVLYNARPDAGLYT 60

Db 345 ARDCOVQGRPEVETWIRNGIEMEL--AKDQKRY---IORGALITLNAQPSDTMTWQ 398
 QY 61 CTARNAGLLRADPFLSVYQ 80
 Db 399 CEARNRHGILLNAYITVYQ 418

RESULT 2
 Q91897

PRELIMINARY; PRT; 814 AA.

AC Q91897;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE FIBROBLAST GROWTH FACTOR RECEPTOR PRECURSOR.
 GN X1FGR OR E59.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Friesel R.E., David I.B.;
 RL Submitted (Aug-1991) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 622-677 FROM N.A.
 RX MEDLINE-95383727; PubMed-7655077;
 RA Brandt A.W., Kirschner M.W.;
 RT "Molecular cloning of tyrosine kinases in the early Xenopus embryo:
 RT identification of Eck-related genes expressed in cranial neural crest
 RT cells of the second (hyoid) arch."
 RT Dev. Dyn. 203:119-140(1995).
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.

DR EMBL; M55163; AAA49990.1; -;
 DR EMBL; U11723; AAA91286.1; -;
 DR HSSP; P08631; IAD5.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam; PF00047; Ig; 3.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR SMART; SM00408; IGC2; 3.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ABP; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Signal; Transferase; Tyrosine-protein kinase.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 81 FIBROBLAST GROWTH FACTOR RECEPTOR.
 SO SEQUENCE 814 AA; 90681 MW; B47CE05F0A8F962A CRC64;

Query Match 28.5%; Score 126; DB 13; Length 814;
 Best Local Similarity 36.2%; Pred. No. 1.8e-06;

Matches 29; Conservative 8; Mismatches 37; Indels 6; Gaps 2;

QY 5 CDVSGRPAPVATWQSHORENLIRPDQMGVNVVTS-IGQVLVYNAREPDAGLYTCR 63
 Db 174 CPNAGFPSPALRWLNKKE-----FRPDQRIIGYKRVRSQWSLIMSVPSPDGNITCIY 228
 QY 64 RNAGLLRADPFLSVYQREP 83
 Db 229 ENRYGTILNHTYLDYVERSP 248

RESULT 3
 Q9BXN7
 ID Q9BXN7 PRELIMINARY; PRT; 497 AA.

AC Q9BXN7;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE FGF HOMOLOGOUS FACTOR RECEPTOR.
 GN FHR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aggarwal S., Xie M.-H., Foster J., Frantz G., Stinson J., Corpuz R.T.,
 RA Simmons L., Hillan K., Yansura D.G., Vandlen R.L., Goddard A.D.,
 RA Gunney A.L.;
 RT "FHR, a novel fibroblast growth factor receptor that uniquely binds
 RT the fibroblast growth factor homologous factors."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF312678; AAK15273.1; -;
 KW Receptor.

SO SEQUENCE 497 AA; 53757 MW; 57301F4F36357360 CRC64;

Query Match 28.4%; Score 125.5; DB 4; Length 497;
 Best Local Similarity 36.7%; Pred. No. 1.2e-06;
 Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;

QY 3 LHCVSGRPAPVATWQSHORENLIRPDQMGVNVVTSIGQVLVYNAREPDAGLYTCR 62
 Db 170 LCVASHPREDITWMD---DQALTRPAAEPR--KKWTLISKNLPEDSGYTCR 222
 QY 63 ARNAGLLRADPFLSVYQ 81
 Db 223 VSNRAGALNATKYVDIQR 241

RESULT 4
 ID Q9HAD7 PRELIMINARY; PRT; 504 AA.
 AC Q9HAD7;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE FGF-LIKE PROTEIN PRECURSOR (FIBROBLAST GROWTH FACTOR RECEPTOR 5).
 GN FGFRL OR FGFR5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CARTILAGE;
 RC PubMed-11031111;
 RA Wiedemann M., Trueb B.;
 RT "Characterization of a novel protein (FGFRL) from human cartilage
 RT related to FGF receptors."
 RT Genomics 69:275-279(2000).
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL; AJ277437; CAC14171.1; -;
 DR EMBL; AF279689; AAK26742.1; -;
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 3.

DR SMART: SM00409: IG: 3.
 DR SMART: SM00408: IGC2: 3.
 DR SMART: SM00410: IG1like: 2.
 KW Signal: Receptor.
 FT SIGNAL 1
 SO SEQUENCE 504 AA: 54567 MW: 16382E57D4276485 CRC64: POTENTIAL.

Query Match 28.4%; Score 125.5; DB 4; Length 504;
 Best Local Similarity 36.7%; Pred. No. 1.2e-06;
 Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;

OY 3 LHCDSGRRPPAVTWKSHORENLIMRPPDMGNYVVSIGQLVLYNARPDAGLYTCT 62
 DB 170 LHCVASGHRPDIITWTKD-----DQALTRPEAAEPR---KKWTLSLKNLRPEDSKYTCR 222
 OY 63 ARNAGLLRADPPLSVYQR 81
 DB 223 VSNRAGAINATYKVDYICR 241

RESULT 5
 ID 090Y07 PRELIMINARY: PRT: 538 AA.
 AC 090Y07:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 (FRAGMENT).
 GN LICAM OR L1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=BRAIN;
 RA Maygar J.P.;
 RL Submitted (DEC-1991) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=BRAIN;
 RA MEDLINE=93021240; PubMed=1404492;
 RX Kohl A., Glese K.P., Mohajeri M.H., Montag D., Moos M., Schachner M.;
 RT "Analysis of promoter activity and 5' genomic structure of the neural
 cell adhesion molecule L1."
 RL J. Neurosci. Res. 32:167-177(1992).
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 DOMAIN.
 DR EMBL: X63510; CAB57301.1; -.
 DR EMBL: X63511; CAB57301.1; JOINED.
 DR EMBL: X63512; CAB57301.1; JOINED.
 DR HSSP: P40223; ICGF.
 DR MGD: MGI:96721; Licam.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_1like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig_5.
 DR SMART: SM00408; IGC2; 4.
 DR SMART: SM00410; IG1like; 1.
 FT NON_TER 538
 SO SEQUENCE 538 AA: 60671 MW: EEOEOC1E235F7B4E CRC64;

Query Match 27.8%; Score 123; DB 11; Length 538;
 Best Local Similarity 38.8%; Pred. No. 2.6e-06;
 Matches 31; Conservative 7; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCDVSGRRPPAVTWKSHORENLIMRPPDMGNYVVSIGQLVLYNARPDAGLYT 60
 DB 349 ARLDQVQGRPOPETITWRINGSMET--VNKDQYR----IEGSLILSNVQPSDTMVTQ 402

OY 61 CTARNAGLLRADPPLSVYQ 80
 DB 403 CEARNQHGILLANAYTYVQ 422

RESULT 6
 ID 090Y38 PRELIMINARY: PRT: 1259 AA.
 AC 090Y38:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1.
 GN ABCD1 OR LICAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Platzer M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
 RA Rosenthal A.;
 RT "Comparative sequence analysis of the mouse Licam locus and the
 corresponding region of human Xq28."
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 DOMAIN.
 DR EMBL: AF135093; AAF22153.1; -.
 DR HSSP: P20241; ICFB.
 DR MGD: MGI:1349215; Abcd1.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_1like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00041; fn3; 4.
 DR Pfam: PF00047; Ig; 6.
 DR SMART: SM00060; FN3; 3.
 DR SMART: SM00408; IGC2; 5.
 DR SMART: SM00410; IG1like; 1.
 SO SEQUENCE 1259 AA: 140916 MW: 25743C039892A22F CRC64;

Query Match 27.8%; Score 123; DB 11; Length 1259;
 Best Local Similarity 38.8%; Pred. No. 7e-06;
 Matches 31; Conservative 7; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCDVSGRRPPAVTWKSHORENLIMRPPDMGNYVVSIGQLVLYNARPDAGLYT 60
 DB 349 ARLDQVQGRPOPETITWRINGSMET--VNKDQYR----IEGSLILSNVQPSDTMVTQ 402
 OY 61 CTARNAGLLRADPPLSVYQ 80
 DB 403 CEARNQHGILLANAYTYVQ 422

RESULT 7
 ID 09CWM1 PRELIMINARY: PRT: 379 AA.
 AC 09CWM1:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 2410003B16RIK PROTEIN.
 GN 2410003B16RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombert P.,
RA Norone P., Rung B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
CC -i- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC EMBL: AK010350; BAB26871.1; -;
DR MGD: MGI:1919583; 2410003B16R1K.
DR InterPro: IPR003599; Ig_
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_c2.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3.
DR SMART: SM00409; Ig_3.
DR SMART: SM00408; Ig_c2; 3.
DR SMART: SM00410; Ig_1like; 1.
SO SEQUENCE 379 AA; 42157 MW; 42A96BEDAAC88F25 CRC64;

Query Match 27.5%; Score 121.5; DB 11; Length 379;
Best Local Similarity 36.2%; Pred. No. 2.6e-06;
Matches 29; Conservative 13; Mismatches 27; Indels 11; Gaps 3;
OY 3 LHCDSGPPPAVWTEKQ---SHORENLMRPDMYGNVVTISIGOLVLYNARPEADGL 58
DB 249 LECRVSGVPPQIFWKKNESLFTSTERSVMHODN-HGYIC-----LLIGATKEDAGW 301
OY 59 YTCARNAAGLLRADPFLSV 78
DB 302 YTVSAKNEAGIVSCTARLDV 321
RESULT 8
O9ERT54 PRELIMINARY; PRT; 492 AA.
AC O9ERT54; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ACTIN-ASSOCIATED PROTEIN PALLADIN (FRAGMENT).
GN 2410003B16R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/NIH;
RX MEDLINE=20391984; PubMed=10931874;
RA Farast M.M., Oley C.A.;
RT "Characterization of palladin, a novel protein localized to stress
RT fibers and cell adhesions.";
RL J. Cell Biol. 150:643-656(2000)
CC -i- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC EMBL: AF205078; AAG00078.1; -;
DR MGD: MGI:1919583; 2410003B16R1K.
DR InterPro: IPR003599; Ig.

DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_1like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3.
DR SMART: SM00409; Ig_3.
DR SMART: SM00408; Ig_c2; 3.
DR SMART: SM00410; Ig_1like; 1.
FT NON TER 1
SO SEQUENCE 492 AA; 55072 MW; BC59E5B3E3BAEBD5 CRC64;
Query Match 27.5%; Score 121.5; DB 11; Length 492;
Best Local Similarity 36.2%; Pred. No. 3.5e-06;
Matches 29; Conservative 13; Mismatches 27; Indels 11; Gaps 3;
OY 3 LHCDSGPPPAVWTEKQ---SHORENLMRPDMYGNVVTISIGOLVLYNARPEADGL 58
DB 362 LECRVSGVPPQIFWKKNESLFTSTERSVMHODN-HGYIC-----LLIGATKEDAGW 414
OY 59 YTCARNAAGLLRADPFLSV 78
DB 415 YTVSAKNEAGIVSCTARLDV 434

RESULT 9
O9NR99 PRELIMINARY; PRT; 2828 AA.
AC O9NR99; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ADICAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA.
RA Crowl R.M., Luk D.;
RT "Identification of the gene encoding Adican, a novel protein
RT expressed in human arthritic tissues.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -i- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC EMBL: AF245505; AAF86402.1; -;
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_1like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_cyp.
DR Pfam: PF00047; Ig_12.
DR Pfam: PF01463; LRRCT; 1.
DR SMART: SM00408; Ig_c2; 10.
DR SMART: SM00410; Ig_1like; 2.
DR SMART: SM00370; LRR; 2.
DR SMART: SM00082; LRRCT; 1.
DR SMART: SM00013; LRRNT; 1.
DR SMART: SM00369; LRR_TYP; 3.
SO SEQUENCE 2828 AA; 312292 MW; A18377D854F1FE1 CRC64;
Query Match 27.5%; Score 121.5; DB 4; Length 2828;
Best Local Similarity 36.7%; Pred. No. 2.7e-05;
Matches 29; Conservative 10; Mismatches 37; Indels 3; Gaps 1;
OY 5 CDVSGPPPAVWTEKQSHORENLMRPDMYGNVVTISIGOLVLYNARPEADGLTGTAR 64
DB 1875 CEATGKRPFFVWTIKVS---TGALMPTNRIDREFVLKNGTLVIRKVVODRGQYMTAS 1931
OY 65 NAAGLRADPFLSVQREP 83

```
DB 1932 NLHGDRMVLSVTVQOP 1950

RESULT 10
043608 PRELIMINARY: PRT: 285 AA.
AC 043608:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ROUNDABOUT 2 (FRAGMENT).
GN ROBO2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors *;
RL Cell 92:205-215(1998).
CC -1- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: AF040991; AAC39576.1; -.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00047; Ig; 2.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00408; IGC2; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 285 AA; 30606 MW; 05DF916A3DBA96C6 CRC64;

Query Match 27.3%; Score 120.5; DB 4; Length 285;
Best Local Similarity 32.3%; Pred. No. 2.4e-06;
Matches 30; Conservative 12; Mismatches 36; Indels 15; Gaps 3;

QY 5 CDVSGRPAPVATWQKSHQRENILM--RPDQMGVNVVTSIGQLVYNARPEDAGLYTCT 62
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13 CERKGNPQPAVFWQKSGSQ--NLFPNQPOQPSRCSVSPTGDLTTNQRSDAGYICO 70
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 ARNAGLLRA-----DEPLSVQREPA 84
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 71 ALTVAGSILAKAQLEVTVLDRPPIILQGPA 103
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
09HCK4 PRELIMINARY: PRT: 1380 AA.
AC 09HCK4:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIA1568 PROTEIN (FRAGMENT).
GN KIA1568.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT Prediction of the coding sequences of unidentified human
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.*;
```

```
RL DNA Res. 7:273-281(2000).
CC -1- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: AB046788; BAB1394.1; -.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003599; Ig_.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 3.
DR SMART: SM00409; IG; 5.
DR SMART: SM00408; IGC2; 5.
DR SMART: SM00406; IGV; 3.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1380 AA; 151426 MW; 45EF63E2EFA26732 CRC64;

Query Match 27.3%; Score 120.5; DB 4; Length 1380;
Best Local Similarity 32.3%; Pred. No. 1.5e-05;
Matches 30; Conservative 12; Mismatches 36; Indels 15; Gaps 3;

QY 5 CDVSGRPAPVATWQKSHQRENILM--RPDQMGVNVVTSIGQLVYNARPEDAGLYTCT 62
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 337 CERKGNPQPAVFWQKSGSQ--NLFPNQPOQPSRCSVSPTGDLTTNQRSDAGYICO 394
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 ARNAGLLRA-----DEPLSVQREPA 84
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 395 ALTVAGSILAKAQLEVTVLDRPPIILQGPA 427
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
Q9QZ13 PRELIMINARY: PRT: 1060 AA.
AC Q9QZ13:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.";
RL Cell 96:795-806(1999).
CC -1- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: AF182037; AAF04558.1; -.
DR HSP: P56276; ITIK.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF00047; Ig; 5.
DR PRINTS: PR00014; FNTYPEIII.
DR SMART: SM00060; FN3; 3.
DR SMART: SM00408; IGC2; 4.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWNW_1.
KW Receptor; Repeat.
FT NON_TER 1060
FT NON_TER 1060
SQ SEQUENCE 1060 AA; 116790 MW; C4BC8C11E8542DA4 CRC64;

Query Match 27.1%; Score 120; DB 11; Length 1060;
```

```

Best Local Similarity 36.8%; Pred. No. 1.3e-05;
Matches 28; Conservative 9; Mismatches 35; Indels 4; Gaps 2;

QY 5 CDVSGRPPAYTWKQSHQRENILIM--RPDQYGNVYTSIGQLVLYNARPEADAGLYCT 62
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 338 CFTKGNPQPAVWQKQESQ--NLTFPNQDPNCSKCSVPTDITITNQRSDAGYITCO 395
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 ANNAAGLRADPLSV 78
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 ALTVAGSLAKALEY 411
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
Q92214 PRELIMINARY; PRT; 1344 AA.
AC Q92214;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RIG-1 PROTEIN.
GN RBIG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan S.-S.F., Cox L.A., Dasika G.K., Lee E.Y.-H.P.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
   DOMAIN.
DR EMBL; AF060570; AAD11628.1; -.
DR HSSP; P56276; 1TLK.
DR MGD; MGI:1343102; Rb1g1.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00408; IGC2; 5.
SQ SEQUENCE 1344 AA; 143439 MW; 8B0060341C49CFEA CRC64;

Query Match 26.98; Score 119; DB 11; Length 1344;
Best Local Similarity 32.5%; Pred. No. 2.2e-05;
Matches 26; Conservative 15; Mismatches 33; Indels 6; Gaps 2;

QY 2 SLHCDVSGRPPAYTWKQSHQRENILIMRPDQY--YGNVYTSIGQLVLYNARPEADAGL 58
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 343 SFQCTKGNPQPAVWQKQESQ--VLLFPQSILQPMGRLLVSPRGQLTITEVKIGDGY 399
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 59 YTCARNAAGLRADPLSV 78
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 YVCAVSAVAGSLAKALLEI 419
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
Q9JUL1 PRELIMINARY; PRT; 1252 AA.
AC Q9JUL1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NEIGHBOR OF PUNC E11 PROTEIN.
GN NOPE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FVB;

```

```

RX MEDLINE=20175427; PubMed=10708514;
RA Saibam J.M., Kappen C.;
RT "Cloning and expression of Nope, a new mouse gene of the
RL immunoglobulin superfamily related to guidance receptors.";
RT Genomics 64:15-23(2000).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
   DOMAIN.
DR EMBL; AF176694; AAF65930.1; -.
DR MGD; MGI:1858497; Nope.
DR InterPro; IPR003962; FN.III_repeat.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; FNYPEPITI.
DR SMART; SM00060; FN3; 5.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; Ig_Like; 1.
KW Repeat.
SQ SEQUENCE 1252 AA; 134759 MW; 11948773277B76B2 CRC64;

Query Match 26.6%; Score 117.5; DB 11; Length 1252;
Best Local Similarity 35.7%; Pred. No. 3.1e-05;
Matches 30; Conservative 11; Mismatches 24; Indels 19; Gaps 3;

QY 5 CDVSGRPPAYTWKQSHQRENILIMRPDQY--GNNVYTSIGQLVLYNARPEADAGL 58
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 355 CRASGEPRPALHW-----LHDGIPLRPNGRVKYGGGGSLVITQIGL-----QDAGY 401
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 59 YTCARNAAGLRADPLSVQORE 82
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 YVCAVSAVAGSLAKADPLAVVRE 425
   1: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
Q9EQS9 PRELIMINARY; PRT; 1252 AA.
AC Q9EQS9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DDM36.
GN DDM36.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/CBA;
RA Murakami H., Nakamata T., Nakayama T., Yamamoto H., Hosaka T.,
RA Aoyama T., Nagayama S., Oka M., Kiyono T., Sasaki M.S., Nakamura T.,
RA Toguchida J.;
RT "Up-regulation of a ras effector and down-regulation of a cell
RT adhesion molecule are associated with transformation of osteoblasts.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
   DOMAIN.
DR EMBL; AB052620; BAB19278.1; -.
DR InterPro; IPR003962; FN.III_repeat.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; FNYPEPITI.
DR SMART; SM00060; FN3; 5.
DR SMART; SM00409; Ig; 4.

```


DR SMART: SM00408; IGC2: 4.
DR SMART: SM00410; IG_Like: 2.
KW Repeat.
SQ SEQUENCE 1252 AA; 134764 MW; BA5292393483AB73 CRC64;

Query Match 26.6%; Score 117.5; DB 11; Length 1252;

Best Local Similarity 35.7%; Pred. NO. 3.1e-05; Mismatches 24; Indels 19; Gaps 3;

Matches 30; Conservative 11; Mismatches 24; Indels 19; Gaps 3;

QY 5 CDVSGRPPPAVWMEKOSHORENLIMRPDMY-----GNVVVTSIGQLVLYNARPEDAGL 58
DB 355 CRASGEPRPALHW-----LHDCIPLRPNGRVRKVGCGGSLVTQTGL-----QDAGY 401

QY 59 YTCIARNAGLLRADEPLSVORE 82

DB 402 YOCVAENSAGTACAAAPLAVVRE 425

Search completed: February 26, 2002, 01:49:20
Job time: 1269 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:38 ; Search time 45.78 seconds
(without alignments)
67.275 Million cell updates/sec

Title: US-09-819-136-2-COPY_203_286
Perfect score: 442
Sequence: 1 ASLHCDVSGRPPAVTWKQ.....NAAGLLRADPPLSVQREPA 84

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	29.0	898	1 FAS2_SCHAM	P22648 schistocerc
2	126.5	28.6	1070	1 PTR7_HUMAN	Q13308 rat musc
3	123.5	27.9	3707	1 PGBM_MOUSE	Q05793 mus muscu
4	122	27.6	1259	1 CAML_RAT	Q05695 rattus norv
5	122	27.6	1260	1 CAML_MOUSE	P11627 mus muscu
6	121.5	27.5	1906	1 KMLS_CHICK	P11799 gallus gall
7	121	27.4	1257	1 CAML_HUMAN	P32004 homo sapien
8	114	25.8	802	1 FGR4_HUMAN	P23455 homo sapien
9	111.5	25.2	1051	1 PTR7_CHICK	Q91048 gallus gall
10	110.5	25.0	1348	1 VGR2_COTJA	P52583 coturnix co
11	110	24.9	812	1 FGR1_XENLA	P22182 xenopus lae
12	109	24.7	801	1 FGR3_MOUSE	Q61851 mus muscu
13	109	24.7	806	1 FGR3_HUMAN	P23607 homo sapien
14	109	24.7	2012	1 DSCA_HUMAN	O60469 homo sapien
15	108.5	24.5	1343	1 VGR2_RAT	Q08775 rattus norv
16	108.5	24.5	1447	1 DCC_MOUSE	P70211 mus muscu
17	107.5	24.3	4393	1 PGBM_HUMAN	P96160 homo sapien
18	107	24.2	822	1 FGR1_RAT	Q04589 rattus norv
19	106	24.0	725	1 NCA2_MOUSE	P13594 mus muscu
20	106	24.0	822	1 FGR1_HUMAN	P11362 homo sapien
21	106	24.0	822	1 FGR1_MOUSE	P16092 mus muscu
22	106	24.0	1115	1 NCA1_MOUSE	P13595 mus muscu
23	105.5	23.9	1367	1 VGR2_MOUSE	P33918 mus muscu
24	105.5	23.9	1447	1 DCC_HUMAN	P43146 homo sapien
25	105	23.8	1914	1 KMLS_HUMAN	Q15746 homo sapien
26	104	23.5	1091	1 NCA1_CHICK	P13590 gallus gall
27	103	23.3	761	1 NCA2_HUMAN	P13592 homo sapien
28	103	23.3	848	1 NCA1_HUMAN	P13591 homo sapien
29	103	23.3	853	1 NCA1_BOVIN	P13836 bos taurus
30	102.5	23.2	537	1 MYPH_CHICK	Q05623 gallus gall
31	102.5	23.2	1142	1 MYPH_HUMAN	Q14324 homo sapien
32	102	23.1	806	1 CER2_CHICK	P18460 gallus gall
33	102	23.1	819	1 FGR1_CHICK	P21804 gallus gall

ALIGNMENTS

34	102	23.1	858	1 NCA1_RAT	P13596 rattus norv
35	100.5	22.7	1336	1 VGR1_RAT	P53767 rattus norv
36	100	22.6	811	1 FAS2_DROME	P34083 drosophila
37	100	22.6	873	1 FAS2_DROME	P34082 drosophila
38	99.5	22.5	1010	1 COMT_CHICK	P14781 gallus gall
39	99.5	22.5	1020	1 COMT_MOUSE	P12960 mus muscu
40	98	22.2	1302	1 NRC_DROME	P20241 drosophila
41	97	21.9	654	1 BFR2_HUMAN	Q01742 homo sapien
42	97	21.9	808	1 FGR4_MOUSE	Q03142 mus muscu
43	97	21.9	821	1 FGR2_HUMAN	P21802 homo sapien
44	96.5	21.8	622	1 VEIN_DROME	Q94918 drosophila
45	96.5	21.8	1147	1 KMLS_RABIT	P29294 oryctolagus

RESULT 1
FAS2_SCHAM
ID FAS2_SCHAM STANDARD: PRT: 898 AA.
AC P22648:
DT 01-AUG-1991 (rel. 19, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE FASCICLIN II PRECURSOR (FAS II).
GN FAS2.
OS Schistocerca americana (American grasshopper).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caellifera;
OC Acridomorpha; Acridoidea; Acrididae; Schistocerca.
OX NCBI_TaxID=7009;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89043938; PubMed=3187519;
RA Harrelson A.L., Goodman C.S.;
RT "Growth cone guidance in insects: fasciclin II is a member of the
RT immunoglobulin superfamily."; Science 242:700-708(1988).
RN [2]
RP SEQUENCE OF 423-436.
RX MEDLINE=88276943; PubMed=2839842;
RA Snow P.M., Zinn K., Harrelson A.L., McAllister L., Schilling J.,
RA Bastiani M.J., Makk G., Goodman C.S.;
RT "Characterization and cloning of fasciclin I and fasciclin II
RT glycoproteins in the grasshopper.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:5291-5295(1988).
CC -1- FUNCTION: NEURONAL RECOGNITION MOLECULE. INVOLVED IN A PATHWAY
CC RECOGNITION FOR AXONS DURING THE DEVELOPMENT OF NERVE FASCICLES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: J03789; AAA29810.1; -
CC PIR: A40114; A40114.
CC PIR: B31817; B31817.
CC InterPro: IPR001777; FN_III.
CC InterPro: IPR003006; IG_MHC.
CC InterPro: IPR003598; IG_C2.
CC InterPro: IPR003600; IG_Like.
CC Pfam: PF00041; fn3_2.
CC Pfam: PF00047; fn3_5.
CC PRINTS: PR00014; ENTPEP111.
CC SMART: SM00060; FN3_2.
CC SMART: SM00408; IGC2_4.
CC SMART: SM00410; IG_Like; 1.

Query Match	29.0%	Score 128;	DB 1;	Length 898;
Best Local Similarity	39.0%	Pred. No. 3.4e-07;		
Matches 30;	Conservative 14;	Mismatches 29;	Indels 4;	Gaps 2;

Db 304 NAAGTATATVEVTVVR 320

receptor protein tyrosine kinase-like molecule closely related to

[illegible]

```

FT CONFLICT 147 147 K -> T (IN REF. 2).
FT CONFLICT 207 207 S -> G (IN REF. 2).
FT CONFLICT 495 496 VL -> RV (IN REF. 2).
FT CONFLICT 515 515 G -> E (IN REF. 2).
FT CONFLICT 834 834 S -> T (IN REF. 2).
FT CONFLICT 881 881 E -> G (IN REF. 2).
FT CONFLICT 969 969 A -> P (IN REF. 2).
FT CONFLICT 992 992 S -> F (IN REF. 2).
SO SEQUENCE 1070 AA: 118260 MW: 47CDF258BE3698A5 CRC64;

Query Match 28.6%; Score 126.5; DB 1; Length 1070;
Best Local Similarity 30.9%; Pred. No. 6,2e-07;
Matches 25; Conservative 17; Mismatches 34; Indels 5; Gaps 1;

OY 1 ASLHCDVGRPPAVTWKSHORENIMRPMQYGVNTSTIGOLVLYNARPEAGLYT 60
DB 609 ALQCEACQDPKPLUQKDR-----ILDPTKLGRMHIFQNGSLVHDVAPEDSGRYT 663
OY 61 CTARNAGLLRADPPLSYQVR 81
DB 664 CIAGNSCNIKTEAPLYVDK 684

RESULT 3
PGBL_MOUSE STANDARD; PRT; 3707 AA.
ID PGBL_MOUSE
AC 005793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R., Valente P., Cal S., Horigan E., Sasaki M.,
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule."
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan."
RL J. Biol. Chem. 263:15379-15387(1988).
CC -1- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC SUBSTRATE FOR CELLS.
CC -1- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC DIMERS OR STELLATE STRUCTURES.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS.
CC -1- SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M77174; AAA39911.1; -
CC EMBL: J04054; AAA39899.1; -
CC EMBL: J04055; AAA39912.1; -
CC HSSP: P01130; IAUJ.
CC MGD: MGI:96257; Hspg2.
CC InterPro: IPR000561; EGF_1-like.
CC InterPro: IPR001438; EGF_1.
CC InterPro: IPR001438; EGF_1.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003598; Ig_c2.
CC InterPro: IPR002172; LDL_recept_A.
CC InterPro: IPR000034; Laminin_B.
CC InterPro: IPR002049; Laminin_Egf.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR000082; SEA.
CC Pfam: PF00047; Ig_15.
CC Pfam: PF00052; Laminin_B_3.
CC Pfam: PF00053; Laminin_EGF.
CC Pfam: PF00054; Laminin_G_3.
CC Pfam: PF00057; Ldl_recept_a; 4.
CC Pfam: PF01390; SEA; 1.
CC PRINTS: PR00010; EGFBLOOD.
CC PRODOM: PD003031; Laminin_B_3.
CC SMART: SM00180; EGF_Lam; 7.
CC SMART: SM00001; EGF_like; 6.
CC SMART: SM00408; ICG2; 14.
CC SMART: SM00281; Lamb; 3.
CC SMART: SM00192; LDLa; 4.
CC SMART: SM00200; SEA; 1.
CC PROSITE: PS00022; EGF_1; 8.
CC PROSITE: PS01186; EGF_2; 5.
CC PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
CC PROSITE: PS01209; LDLa_1; 4.
CC PROSITE: PS0068; LDLa_2; 4.
CC PROSITE: PS00024; SEA; 1.
CC K1: Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
CC K2: Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
CC K3: Extracellular matrix; EGF-like domain.
CC SIGNAL 1 21
CC CHAIN 22 3707
FT FT 80 194
FT FT 195 234
FT FT 281 319
FT FT 320 359
FT FT 360 403
FT FT 404 504
FT FT 521 530
FT FT 531 730
FT FT 731 763
FT FT 764 813
FT FT 814 871
FT FT 879 923
FT FT 924 933
FT FT 934 1125
FT FT 1126 1158
FT FT 1159 1208
FT FT 1209 1265
FT FT 1275 1324
FT FT 1325 1334
FT DOMAIN 1334

```

FT DOMAIN 1335 1529 LAMININ DOMAIN IV 3 (DOMAIN III C).
 FT DOMAIN 1562 1562 LAMININ EGF-LIKE 9 (C-TERMINAL).
 FT DOMAIN 1563 1612 LAMININ EGF-LIKE 10.
 FT DOMAIN 1613 1670 LAMININ EGF-LIKE 11.
 FT DOMAIN 1677 1771 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 1772 1865 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 1866 1954 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 1955 2049 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 2050 2148 IG-LIKE C2-TYPE DOMAIN 6.
 FT DOMAIN 2149 2244 IG-LIKE C2-TYPE DOMAIN 7.
 FT DOMAIN 2245 2343 IG-LIKE C2-TYPE DOMAIN 8.
 FT DOMAIN 2344 2436 IG-LIKE C2-TYPE DOMAIN 9.
 FT DOMAIN 2437 2532 IG-LIKE C2-TYPE DOMAIN 10.
 FT DOMAIN 2533 2619 IG-LIKE C2-TYPE DOMAIN 11.
 FT DOMAIN 2620 2720 IG-LIKE C2-TYPE DOMAIN 12.
 FT DOMAIN 2721 2809 IG-LIKE C2-TYPE DOMAIN 13.
 FT DOMAIN 2810 2895 IG-LIKE C2-TYPE DOMAIN 14.
 FT DOMAIN 2896 2980 IG-LIKE C2-TYPE DOMAIN 15.
 FT DOMAIN 2981 3130 LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
 FT DOMAIN 3049 3241 EGF-LIKE 1.
 FT DOMAIN 3241 3495 EGF-LIKE 2.
 FT DOMAIN 3304 3495 LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
 FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
 FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
 FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
 FT SITE 3615 3617 MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 206 225 BY SIMILARITY.
 FT DISULFID 219 234 BY SIMILARITY.
 FT DISULFID 285 297 BY SIMILARITY.
 FT DISULFID 292 310 BY SIMILARITY.
 FT DISULFID 304 319 BY SIMILARITY.
 FT DISULFID 325 337 BY SIMILARITY.
 FT DISULFID 332 350 BY SIMILARITY.
 FT DISULFID 344 359 BY SIMILARITY.
 FT DISULFID 368 381 BY SIMILARITY.
 FT DISULFID 375 394 BY SIMILARITY.
 FT DISULFID 388 403 BY SIMILARITY.
 FT DISULFID 428 479 BY SIMILARITY.
 FT DISULFID 764 773 BY SIMILARITY.
 FT DISULFID 766 780 BY SIMILARITY.
 FT DISULFID 783 792 BY SIMILARITY.
 FT DISULFID 795 811 BY SIMILARITY.
 FT DISULFID 814 829 BY SIMILARITY.
 FT DISULFID 816 839 BY SIMILARITY.
 FT DISULFID 842 851 BY SIMILARITY.
 FT DISULFID 854 869 BY SIMILARITY.
 FT DISULFID 1159 1168 BY SIMILARITY.
 FT DISULFID 1161 1175 BY SIMILARITY.
 FT DISULFID 1178 1187 BY SIMILARITY.
 FT DISULFID 1190 1206 BY SIMILARITY.
 FT DISULFID 1209 1224 BY SIMILARITY.
 FT DISULFID 1211 1234 BY SIMILARITY.
 FT DISULFID 1237 1246 BY SIMILARITY.
 FT DISULFID 1249 1263 BY SIMILARITY.
 FT DISULFID 1275 1287 BY SIMILARITY.
 FT DISULFID 1277 1293 BY SIMILARITY.
 FT DISULFID 1295 1304 BY SIMILARITY.
 FT DISULFID 1307 1322 BY SIMILARITY.
 FT DISULFID 1563 1572 BY SIMILARITY.
 FT DISULFID 1565 1579 BY SIMILARITY.
 FT DISULFID 1582 1591 BY SIMILARITY.
 FT DISULFID 1594 1610 BY SIMILARITY.
 FT DISULFID 1613 1628 BY SIMILARITY.
 FT DISULFID 1615 1638 BY SIMILARITY.
 FT DISULFID 1641 1650 BY SIMILARITY.
 FT DISULFID 1653 1668 BY SIMILARITY.
 FT DISULFID 1792 1839 BY SIMILARITY.
 FT DISULFID 1886 1932 BY SIMILARITY.
 FT DISULFID 1976 2021 BY SIMILARITY.
 FT DISULFID 2073 2118 BY SIMILARITY.
 FT DISULFID 2170 2215 BY SIMILARITY.

FT DISULFID 2268 2313 BY SIMILARITY.
 FT DISULFID 2365 2413 BY SIMILARITY.
 FT DISULFID 2456 2506 BY SIMILARITY.
 FT DISULFID 2554 2599 BY SIMILARITY.
 FT DISULFID 2641 2686 BY SIMILARITY.
 FT DISULFID 2831 2876 BY SIMILARITY.
 FT DISULFID 2917 2962 BY SIMILARITY.
 Query Match 27.9% Score 123.5; DB 1; Length 3707;
 Best Local Similarity 39.0%; Pred. No. 6e-06;
 Matches 30; Conservative 7; Mismatches 29; Indels 11; Gaps 2;
 QY 5 CDVSGRPAPVATWKEQSHRENLIMRPDMYGNVVTSGQLVITYNARPEAGLYCTAR 64
 DB 2917 CMASGYPPIATWKSVDSD-----LPPSRLEN-----NMILPVSVPEDAGTYVCTAT 2965
 QY 65 NAAGILRADPLSVQVR 81
 DB 2966 NROGKVKAFAYLOVPER 2982
 RESULT 4
 CAML_RAT STANDARD; PRT; 1259 AA.
 ID CAML_RAT
 AC Q05695;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GN LICAM OR CAML1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91372414; PubMed=1894011;
 RA Miura M., Kobayashi M., Asou H., Uyemura K.;
 RT "Molecular cloning of cDNA encoding the rat neural cell adhesion
 molecule L1. Two L1 isoforms in the cytoplasmic region are produced
 by differential splicing.";
 RL FEBS Lett. 289:91-95(1991).
 CC - FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO AXONIN ON NEURONS.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC - ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING.
 CC - TISSUE SPECIFICITY: THE SHORTER ISOFORM IS PREDOMINANTLY FOUND IN
 CC THE BRAIN, WHILE THE LONGER ISOFORM IS FOUND IN THE PERIPHERAL
 CC NERVOUS SYSTEM.
 CC - SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC - SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X59149; CAA1860.1; -
 DR PIR: S17655; S17655.
 DR HSSP: P20241; ICFB.
 DR InterPro: IPR001777; FN_III.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR Pfam: PF00041; fn3; 4.
 DR Pfam: PF00047; Ig; 6.
 DR PRINTS: PR00014; FNTYPEP11.


```

FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1180 1183 MISSING (IN SHORT ISOFORM)
FT VARSPLIC 1180 1183 (BY SIMILARITY).
SO SEQUENCE 1260 AA; 140966 MW; 22B57001CB2A538 CRC64;

Query Match 27.6%; Score 122; DB 1; Length 1260;
Best Local Similarity 38.8%; Pred. No. 2.6e-06;
Matches 31; Conservative 7; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCDVSRPPPAYWEQSHORENLINRPMQGTGVVTSIGOLVLYNAPEDAGLYT 60
DB 349 ARLDQVQGRPPQPEITWRINGSMET--VNKQKYR----IQGSLISNVOPTDWTQ 402
OY 61 CTARNAGGLRADPFLSVYQ 80
DB 403 CEARNQHGILLNAYITYVQ 422

RESULT 6
KMLS_CHICK STANDARD; PRT: 1906 AA.
ID KMLS_CHICK STANDARD; PRT: 1906 AA.
AC P11799; P19038;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES
DE (EC 2.7.1.117) (MCKC) [CONTAINS: TELOKIN].
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (MCKC-210).
RX MEDLINE=96033976; PubMed=7589469;
RA Waterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,
RA Stepanova O.V., Shlitsky V.P.;
RT "Multiple gene products are produced from a novel protein kinase
RL transcription region."
RL FEBS Lett. 373:217-220(1995).
RN [2]
RP SEQUENCE FROM N.A. (MCKC-108).
RX MEDLINE=90192792; PubMed=2313320;
RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
RA Means A.R.;
RT "Regulatory and structural motifs of chicken gizzard myosin light
RL chain kinase."
RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
RN [3]
RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Fibroblast;
RX MEDLINE=90361738; PubMed=2202734;
RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,
RA Matrisian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,
RA van Eldik L.J., Waterson D.M.;
RT "Use of DNA sequence and mutant analyses and antisense
RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
RT myosin light chain kinase autoinhibition, calmodulin recognition, and
RT activity."
RT J. Cell Biol. 111:1107-1125(1990).
RN [4]
RP SEQUENCE OF 1259-1906 FROM N.A.
RC TISSUE-Gizzard;
RX MEDLINE=87157387; PubMed=3030394;
RA Guerriero V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.;
RT "Domain organization of chicken gizzard myosin light chain kinase
RL deduced from a cloned cDNA."
RL Biochemistry 25:8372-8381(1986).
RN [5]
RP SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).

```

```

RC TISSUE-Gizzard;
RX MEDLINE=93073972; PubMed=1444462;
RA Yoshikawa S., Ikebe M.;
RT "Molecular cloning of the chicken gizzard telokin gene and cDNA."
RL Arch. Biochem. Biophys. 299:242-247(1992).
RN [6]
RP SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
RX MEDLINE=92236611; PubMed=1373815;
RA Collinge M., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
RA van Eldik L.J., Waterson D.M.;
RT "Structure and expression of a calcium-binding protein gene contained
RT within a calmodulin-regulated protein kinase gene."
RL Mol. Cell. Biol. 12:2359-2371(1992).
CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A
CC MYOSIN LIGHT CHAIN. WHICH LEADS TO THE FORMATION CALMODULIN/
CC MCKC SIGNAL TRANSDUCTION COMPLEXES WHICH ALLOW SELECTIVE
CC TRANSDUCTION OF CALCIUM SIGNALS.
CC -1- CATALYTIC ACTIVITY: ATP + [MYOSIN LIGHT-CHAIN] = ADP + [MYOSIN
CC LIGHT-CHAIN] PHOSPHATE.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST THREE ISOFORMS OF MCKC ARE PRODUCED
CC BY ALTERNATIVE INITIATION, A NON-MUSCLE FORM (MCKC-210) WHICH IS
CC THE LONGEST FORM; A SMOOTH-MUSCLE FORM (MCKC-108) AND TELOKIN (A
CC C-TERMINAL SECTION WITH NO CATALYTIC ACTIVITY).
CC -1- TISSUE SPECIFICITY: TELOKIN IS EXPRESSED IN GIZZARD, HEART, LUNG,
CC INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF THE
CC EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE GIZZARD.
CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52876; CA337056.1; -
DR EMBL: X52876; CA337057.1; -
DR EMBL: X52876; CA337058.1; -
DR EMBL: X51046; AAA49069.1; -
DR EMBL: M14953; AAA69964.1; -
DR EMBL: M96655; AAA49083.1; -
DR EMBL: M88283; AAA48647.1; -
DR EMBL: M88284; AAA53768.1; -
DR PIR: A25810; A25810.
DR PIR: A37099; A37099.
DR PIR: S11652; S11652.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001777; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF00047; fn3; 1.
DR Pfam: PF00069; pkinase; 1.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00408; IGC2; 8.
DR SMART: SM00410; IG_Like; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative Initiation.
FT CHAIN 1 1906 MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
FT FT ISOZYME (MCKC-210).
FT FT MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
FT FT ISOZYME (MCKC-108).

```



```

FT CHAIN 1750 1906 TELOKIN.
FT INT_MET 935 935 FOR MLCK-108.
FT INT_MET 1750 1750 FOR TELOKIN.
FT DOMAIN 1330 1400 FIBRONECTIN TYPE-III.
FT DOMAIN 1453 1708 PROTEIN KINASE.
FT DOMAIN 1809 1876 IG-LIKE C2-TYPE DOMAIN.
FT NP_BIND 1459 1467 ATP (BY SIMILARITY).
FT BINDING 1482 1482 ATP (BY SIMILARITY).
FT ACT_SITE 1574 1574 BY SIMILARITY.
FT DOMAIN 1716 1728 CALMODULIN AUTOINHIBITION (AM13) REGION
(POTENTIAL).
FT DOMAIN 1730 1749 CALMODULIN RECOGNITION (RS20) REGION
(POTENTIAL).
FT DOMAIN 1317 1364 MOTIF IB.
FT DOMAIN 1385 1402 MOTIF IA.
FT REPEAT 660 1833 4 X REPEATS, MOTIF IIA.
FT REPEAT 660 676 IIA-1.
FT REPEAT 758 774 IIA-2.
FT REPEAT 1107 1123 IIA-3.
FT REPEAT 1817 1833 IIA-4.
FT DOMAIN 683 1866 5 X REPEATS, MOTIF IIB.
FT REPEAT 693 708 IIB-1.
FT REPEAT 791 807 IIB-2.
FT REPEAT 1140 1156 IIB-3.
FT REPEAT 1281 1297 IIB-4.
FT REPEAT 1851 1866 IIB-5.
FT DOMAIN 970 1226 4 X REPEATS, MOTIF IIT.
FT REPEAT 970 987 IIT-1.
FT REPEAT 999 1016 IIT-2.
FT REPEAT 1061 1078 IIT-3.
FT REPEAT 1209 1226 IIT-4.
FT DOMAIN 1700 1763 CALMODULIN-BINDING.
FT DOMAIN 1896 1906 POLY-GLU.
FT MOD_RES 1748 1748 PHOSPHORYLATION.
FT MOD_RES 1762 1762 PHOSPHORYLATION.
FT CONFLICT 1439 1439 R -> Q (IN REF. 4).
SQ SEQUENCE 1906 AA; 210445 MW; AD7D8A3B69EE3363 CRC64;

```

```

Query Match 27.5%; Score 121.5; DB 1; Length 1906;
Best Local Similarity 35.4%; Pred. No. 4.7e-06;
Matches 28; Conservative 11; Mismatches 25; Indels 15; Gaps 2;

QY 5 CDVSGRPAPVATWTK-----OSHOENLIMRPDQNGVVSIGQLVLYNRPEDAGLY 59
DB 177 CRIETRPQPOVYTKGDHLLQONERPMFEK-----TGIQYLETONVOLADAGLY 226
QY 60 TCTARNAGLLRADFPLSV 78
DB 227 TCTVNSAGKASVSAELTV 245

```

```

RESULT 7
CAML_HUMAN STANDARD: PRI: 1257 AA.
ID CAML_HUMAN
AC P32004.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1 OR MIC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eulalia; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92031698; PubMed-1932117;
RA Kobayashi M., Miura M., Asou H., Ujemura K.;
RT "Molecular cloning of cell adhesion molecule L1 from human nervous
RT tissue: a comparison of the primary sequences of L1 molecules of
RT different origin.";
RL Biochim. Biophys. Acta 1090:238-240(1991).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Rosenthal A., Couteille O., Drescher B.;
RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-92329299; PubMed-1627459;
RA Reid R.A., Hemperly J.J.;
RT "Variants of human L1 cell adhesion molecule arise through alternate
RT splicing of RNA.";
RL J. Mol. Neurosci. 3:127-135(1992).
RN [4]
RP SEQUENCE OF 353-1176 FROM N.A.
RX MEDLINE-92020233; PubMed-1923824;
RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
RT "PCR walking from microdissection clone M54 identifies three exons
RT from the human gene for the neural cell adhesion molecule L1
RT (CAM-L1).";
RL Nucleic Acids Res. 19:5395-5401(1991).
RN [5]
RP SEQUENCE OF 332-371 FROM N.A.
RX MEDLINE-90353957; PubMed-2387585;
RA Djabali M., Mattei M.-G., Nguyen C., Roux D., Demengeot J.,
RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
RT "The gene encoding L1, a neural adhesion molecule of the
RT immunoglobulin family, is located on the X chromosome in mouse and
RT man.";
RL Genomics 7:587-593(1990).
RN [6]
RP SEQUENCE OF 1030-1257 FROM N.A.
RX MEDLINE-91132183; PubMed-1993895;
RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
RA Stallcup W.B.;
RT "Isolation and sequence of partial cDNA clones of human L1: homology
RT of human and rodent L1 in the cytoplasmic region.";
RL J. Neurochem. 56:797-804(1991).
RN [7]
RP SEQUENCE OF 20-36.
RX MEDLINE-88298876; PubMed-3136168;
RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
RA Rathjen F.G.;
RT "A human brain glycoprotein related to the mouse cell adhesion
RT molecule L1.";
RL J. Biol. Chem. 263:11943-11947(1988).
RN [8]
RP VARIANTS HSAS TYR-264.
RX MEDLINE-94004956; PubMed-8401576;
RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donna D.;
RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus
RT (HSAS).";
RL Nat. Genet. 4:331-331(1993).
RN [9]
RP VARIANTS HSAS/MASA LEU-1194.
RX MEDLINE-95187172; PubMed-7881431;
RA Fransen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
RA Willems P.J.;
RT "X-linked hydrocephalus and MASA syndrome present in one family are
RT due to a single missense mutation in exon 28 of the L1CAM gene.";
RL Hum. Mol. Genet. 3:2255-2256(1994).
RN [10]
RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
RX MEDLINE-95004608; PubMed-7920659;
RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
RA Paterson J., Metzberg A., Ionasescu V., Temple K., Kenwick S.;
RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked
RT hydrocephalus result from mutations in the L1 gene.";
RL Nat. Genet. 7:402-407(1994).
RN [11]
RP VARIANTS MASA GLN-210 AND ASN-598.
RX MEDLINE-95004609; PubMed-7920660;
RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
RA Reviers E., Korn B., Poustka A., Milson G., Schrander-Stumpel C.,
RA Winter R.M., Schwartz C., Willems P.J.;

```

"MASA syndrome is due to mutations in the neural cell adhesion gene L1CAM.";
 RL Nat. Genet. 7:408-413(1994).
 RN [12]
 RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
 RX MEDLINE=95282776; PubMed=7762552;
 RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
 RT Holmberg E., Wadelius C., Kenwick S.;
 RL "New domains of neural cell-adhesion molecule L1 implicated in
 X-linked hydrocephalus and MASA syndrome.";
 Am. J. Hum. Genet. 56:1304-1314(1995).
 RN [13]
 RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
 RX MEDLINE=96153146; PubMed=8556302;
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
 RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
 RT due to mutations in one single gene, L1.";
 Eur. J. Hum. Genet. 3:273-284(1995).
 RN [14]
 RP ERRATUM.
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RL Eur. J. Hum. Genet. 4:126-126(1996).
 RN [15]
 RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
 RX MEDLINE=96057511; PubMed=7562969;
 RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
 RT "Mutations in L1-CAM in two families with X linked complicated
 RT spastic paraplegia, MASA syndrome, and HSAS.";
 J. Med. Genet. 32:549-552(1995).
 RN [16]
 RP VARIANTS HSAS CYS-194 AND LEU-240.
 RX MEDLINE=97083370; PubMed=8929944;
 RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
 RT Engel W., Schwinger E., Gal A.;
 RT "Five novel mutations in the L1CAM gene in families with X linked
 RT hydrocephalus.";
 J. Med. Genet. 33:103-106(1996).
 RN [17]
 RP VARIANTS HSAS O-184; V-439-T-443 DEL; C-784 AND L-936-L-948 DEL.
 RX MEDLINE=97338664; PubMed=9192224;
 RA Macfaiiane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,
 RA Charlton R., Garrett C., Toulme J., Yates J.R.W., Berry C., Goudie D.,
 RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenrick S.;
 RT "Nine novel L1 CAM mutations in families with X-linked
 RT hydrocephalus.";
 Hum. Mutat. 9:512-518(1997).
 RN [18]
 RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
 RX MEDLINE=98180721; PubMed=9521424;
 RA Du Y.-Z., Srivastava A.K., Schwartz C.B.;
 RT "Multiple exon screening using restriction endonuclease
 RT fingerprinting (REF) detection of six novel mutations in the L1 cell
 RT adhesion molecule (L1CAM) gene.";
 Hum. Mutat. 11:222-230(1998).
 RN [19]
 RP VARIANTS CRASH PRO-632.
 RX MEDLINE=98112489; PubMed=9452110;
 RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
 RA Willems P.J.;
 RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
 Hum. Mutat. Suppl. 1:S284-S287(1998).
 RN [20]
 RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
 RX MEDLINE=98415726; PubMed=9744477;
 RA Saugier-Veber F., Marlin C., le Meur N., Lyonnet S., Munnich A.,
 RA David A., Henocq A., Heron D., Jonveaux P., Odent S., Manouvrier S.,
 RA Moncla A., Morichon N., Philip N., Salge D., Tosi M., Frebourg T.;
 RT "Identification of novel L1CAM mutations using fluorescence assisted
 RT mismatch analysis.";
 Hum. Mutat. 12:259-266(1998).
 RL CC
 CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE

CC DEVELOPMENT OF THE NERVOUS SYSTEM, INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO AXONIN ON NEURONS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING.
 CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF THREE X-LINKED
 CC SYNDROMES. 1: HYDROCEPHALUS OWING TO STENOSIS OF THE AQUEDUCT OF
 CC SYLVIUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
 CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
 CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBS.
 CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
 CC MILDER PRESENTATION AND A LONGER LIFE EXPECTANCY. 3: SPASTIC
 CC PARAPLEGIA TYPE 1 (SPG1). COLLECTIVELY THESE SYNDROMES ARE ALSO
 CC KNOWN AS CRASH SYNDROME, AN ACROSYM WHICH STANDS FOR CORPUS
 CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBS,
 CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
 CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF HIRSCHSPRUNG DISEASE
 CC (HSCR).
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- DATABASE: NAME=L1CAM; NOTE=L1CAM mutation Web Page;
 CC WWW="http://hgins.uia.ac.be/dnalab/l1".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X59847; CAA42508.1; -
 CC DR EMBL: Z29373; CAA82564.1; -
 CC DR EMBL: M74387; AAA59476.1; -
 CC DR

Query Match 27.4%; Score 121; DB 1; Length 1257;
 Best Local Similarity 41.2%; Pred. No. 3; 3e-06;
 Matches 33; Conservative 5; Mismatches 36; Indels 6; Gaps 2;

QY 1 ASLHCYSGPPPAVWEKSHORENIMKPDQMGVNVVTSIGLVYNARPDAGLYT 60
 Db 350 ARDCQVQGPQPEVWRINGIVPEEL--AKDKYR-----TORGALITSNVQPSPTMTVQ 403
 QY 61 CTARNAGLLRADEPLSYVQ 80
 Db 404 CEARNRGGILLNAYTYVQ 423

RESULT 8
 ID FGR4_HUMAN STANDARD; PRT; 802 AA.
 AC P22455; Q14309; Q43785;
 DT 01-AUG-1991 (rel. 19) Created)
 DT 20-AUG-2001 (rel. 40, last sequence update)
 DT 20-AUG-2001 (rel. 40, last annotation update)
 DE FIBROBLAST GROWTH FACTOR RECEPTOR 4 PRECURSOR (EC 2.7.1.112) (FGR-4).
 GN FGR4 OR JTK2 OR TKF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91224085; PubMed=1709094;
 RA Partanen J.M., Meckelae T.P., Eerola E., Korhonen J., Hironen H.,
 RA Claesson-Welsh L., Alitalo K.;
 RT "FGR-4, a novel acidic fibroblast growth factor receptor with a
 RT distinct expression pattern.";
 RL EMBD J. 10:1347-1354(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;

RX MEDLINE=93194827; PubMed=7680645;
 RA Ron D., Reich R., Chedid M., Lengel C., Cohen O.E., Chan A.M.,
 RA Neufeld G., Miki T., Tironick S.R.,
 RT "Fibroblast growth factor receptor 4 is a high affinity receptor for
 RT both acidic and basic fibroblast growth factor but not for
 RT keratinocyte growth factor.";
 RL J. Biol. Chem. 268:5388-5394(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98119018; PubMed=9457674;
 RA Kostreva M., Muller U.,
 RT "Genomic structure and complete sequence of the human FGFR4 gene,";
 RL Mamm. Genome 9:131-135(1998).
 RN [4]
 RP SEQUENCE OF 609-676 FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=91062289; PubMed=2247464;
 RA Partanen J., Maekela T.P., Alltalo R., Lehtvaesalho H., Alltalo K.,
 RT "Putative tyrosine kinases expressed in K-562 human leukemia cells,";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8913-8917(1990).
 CC -1- FUNCTION: RECEPTOR FOR ACIDIC FIBROBLAST GROWTH FACTOR. DOES NOT
 CC BIND TO BASIC FIBROBLAST GROWTH FACTOR. BINDS FGFR19.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 CC TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X57205; CAA0490.1; -
 DR EMBL: L03840; AAB59389.1; -
 DR EMBL: Y13901; CAA74200.1; -
 DR EMBL: M59373; AAB63208.1; -
 DR PIR: S15345; TVHUF4.
 DR HSSP: P11362; 1FC1.
 DR MIM: 134935; -
 DR InterPro: IPR003599; I9.
 DR InterPro: IPR003598; I9_c2.
 DR InterPro: IPR003600; I9_1like.
 DR InterPro: IPR002290; Set_thr_kin_actsite.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00047; Ig_3.
 DR Pfam: PF00069; Pkinase; 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR SMART: SM00409; IG; 3.
 DR SMART: SM00408; IGC2; 3.
 DR SMART: SM00410; IG_1like; 2.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TYKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 KW Receptor: Glycoprotein; Tyrosine-protein kinase; ATP-binding;
 KW Transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;
 KW Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 802 FIBROBLAST GROWTH FACTOR RECEPTOR 4.
 FT DOMAIN 25 369 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 370 390 POTENTIAL.
 FT DOMAIN 391 802 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 50 108 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 165 231 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 264 340 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 467 755 PROTEIN KINASE.
 FT NP_BIND 473 481 ATP (BY SIMILARITY).

FT BINDING 503 503 ATP (BY SIMILARITY).
 FT ACT_SITE 612 612 BY SIMILARITY.
 FT MOD_RES 643 643 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT DISULFID 57 101 POTENTIAL.
 FT DISULFID 172 224 POTENTIAL.
 FT CARBOHYD 271 333 POTENTIAL.
 FT CARBOHYD 112 112 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 311 311 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CONFLICT 297 297 D -> V (IN REF. 1).
 SQ SEQUENCE 802 AA; 87954 MW; B22B259831BB889F CRC64;

 Query Match 25.8%; Score 114; DB 1; Length 802;
 Best Local Similarity 32.5%; Pred. No. 1.3e-05;
 Matches 27; Conservative 14; Mismatches 30; Indels 12; Gaps 3;

 QY 5 CTVSGRPAAVYWER--QSHQRENLI--MRPQMGNVYVTSIGQLVLYNAPEDAGLYT 60
 DB 172 CPAAGNPPTIWLMDGQAFHGENRIGIRLRHSHVLSVESV-----VPSDRGYT 223
 QY 61 CTARNAAGLRADFPVYOREP 83
 DB 224 CLVENAVGSRINYLLDYVERSP 246

 RESULT 9
 ID PKT7_CHICK STANDARD; PRT; 1051 AA.
 AC 091048;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TYROSINE-PROTEIN KINASE-LIKE 7 PRECURSOR (KINASE LIKE PROTEIN).
 GN PKT7 OR K1G.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryonic brain;
 RX MEDLINE=91271300; PubMed=1711213;
 RA Chou Y.-H., Hayman M.J.,
 RT "Characterization of a member of the immunoglobulin gene superfamily
 RT that possibly represents an additional class of growth factor
 RT receptor,";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4897-4901(1991).
 CC -1- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
 CC THE CATALYTIC ACTIVITY OF TYROSINE KINASE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW, SPLEEN, BURSA,
 CC THYMUS AND BRAIN. WEAKLY EXPRESSED IN FIBROBLASTS. ALSO EXPRESSED
 CC IN EMBRYONIC LIVER.
 CC -1- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES. TRK-TYPE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63437; AAA48933.1; -
 DR HSSP: P00523; 2PKT.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR003006; Ig_MHC.

DR InterPro: IPR003598; Ig-C2.
 DR InterPro: IPR003600; Ig-like.
 DR InterPro: IPR01245; Tyr_kin.
 DR Pfam: PF00047; Ig_7.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR SMART: SM00408; Igc2; 4.
 DR SMART: SM00410; Ig-like; 4.
 DR SMART: SM00219; TYRC; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_DOM; 1.
 DR Receptor: Transmembrane; Signal; Glycoprotein; Cell adhesion;
 KW Immunoglobulin domain; Repeat
 FT SIGNAL 1 22
 FT CHAIN 23 1051
 FT DOMAIN 23 685
 FT TRANSMEM 686 706
 FT DOMAIN 707 1051
 FT DOMAIN 33 95
 FT DOMAIN 130 194
 FT DOMAIN 227 289
 FT DOMAIN 319 379
 FT DOMAIN 407 469
 FT DOMAIN 498 558
 FT DOMAIN 587 652
 FT DOMAIN 777 1048
 FT DISULFID 40 88
 FT DISULFID 137 187
 FT DISULFID 234 282
 FT DISULFID 326 372
 FT DISULFID 414 462
 FT DISULFID 505 551
 FT DISULFID 594 645
 FT CARBOHD 103 103
 FT CARBOHD 202 202
 FT CARBOHD 255 255
 FT CARBOHD 264 264
 FT CARBOHD 444 444
 FT CARBOHD 548 548
 FT CARBOHD 627 627
 FT CARBOHD 627 627
 FT SEQUENCE 1051 AA; 116366 MW; 1752442AEACB702 CRC64;

Query Match 25.2%; Score 111.5; DB 1; Length 1051;
 Best Local Similarity 28.6%; Pred. No. 3.6e-05;
 Matches 24; Conservative 16; Mismatches 39; Indels 5; Gaps 1;

Oy 1 ASLHCDVSGRPAAVWMEKQSHRENTIMPDQMGVNVTSIGQLVLYNARPEADAGLYT 60
 Db 590 AMFOCOAEGDPVPHIQW-----KGKDKILDSKILPRQIMPNSLIVYDTTDSKRYT 644
 Oy 61 CTARNAGILRADPPLSYVQREPA 84
 Db 645 CIAGNSCNIRKREAFLYVVKPAA 668

RESULT 10
 VGR2_COTJA STANDARD; PRT; 1348 AA.
 AC P52583;
 DT 01-OCT-1996 (Rel. 34, created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 PRECURSOR (EC 2.7.1.112)
 DE (VEGFR-2) (ENDOTHELIAL KINASE RECEPTOR EKI) (OUER 1) (OUER1).
 GN KDR OR FLK-1 OR EKI.
 OS Coturnix coturnix japonica (Japanese quail).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Coturnix.
 OC NCBI_TaxID=93934;
 RN [1]
 KW SEQUENCE FROM N.A.

RC TISSUE=Embryo;
 RX MEDLINE=97017121; PubMed=8863722;
 RA Eichmann A., Marcelle C., Breant C., Le Douarin N.M.;
 RT "Molecular cloning of quak 1 and 2, two quail vascular endothelial
 RL growth factor (VEGF) receptor-like molecules.";
 RN Gene 174:3-8(1996).
 [2]
 RP SEQUENCE OF 910-1348 FROM N.A.
 RC TISSUE=Spinal cord;
 RX MEDLINE=93378866; PubMed=8396413;
 RA Eichmann A., Marcelle C., Breant C., Le Douarin N.M.;
 RT "Two molecules related to the VEGF receptor are expressed in early
 RL endothelial cells during avian embryonic development.";
 RN Mech. Dev. 42:33-48(1993).
 [3]
 RP SEQUENCE OF 764-880 FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Embryo;
 RX MEDLINE=95301109; PubMed=7781909;
 RA Flamme I., Breier G., Risau W.;
 RT "Vascular endothelial growth factor (VEGF) and VEGF receptor 2 (Flk-1)
 RL are expressed during vasculogenesis and vascular differentiation in
 Dev. Biol. 169:699-712(1995).
 CC -1- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN
 CC KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM
 CC PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF
 CC VASCULAR PERMEABILITY.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 CC TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: IN ALL ENDOTHELIAL TISSUES DURING ONSET OF
 CC VASCULARIZATION. IN LATER DEVELOPMENT, PRESENT IN LUNG, HEART,
 CC INTESTINE AND SKIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE MESODERM AT ONSET OF
 CC GASTRULATION. FROM DAY 2, CONFINED TO ENDOTHELIAL TISSUES AND
 CC EXPRESSION CONTINUES TO BE WIDESPREAD THROUGHOUT VASCULARIZATION
 CC UNTIL E9 WHERE IT BECOMES RESTRICTED TO SPECIFIC REGIONS SUCH AS
 CC THE SPINAL CHORD AND HEART VALVES.
 CC -1- INDUCTION: IN VITRO, VEGF IS INDUCED BY BASIC FIBROBLAST GROWTH
 CC FACTOR (FGF), UNIQUELY IN THE FIRST 24 H OF CELL CULTURE.
 CC -1- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isd-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X83288; CAA58268.1; -
 DR EMBL: S65205; AAB28127.1; -
 DR EMBL: S78345; AAB34594.1; -
 DR HSP: P11362; IAGW.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig-like.
 DR InterPro: IPR001824; Receptor_Tyr_kin_III.
 DR Pfam: PF00047; Ig_6.
 DR Pfam: PF00069; pkinase; 2.
 DR SMART: SM00408; Igc2; 1.
 DR SMART: SM00410; Ig-like; 4.
 DR SMART: SM00219; TYRC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00240; RECEPTOR_TYR_KIN_III; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Signal; Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
 KW Glycoprotein; Phosphorylation; ATP-binding.

```
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1348 VASCULAR ENDOTHELIAL GROWTH FACTOR
FT DOMAIN 21 756 RECEPTOR 2.
FT TRANSMEM 777 777 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 778 1348 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 43 106 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 138 202 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 234 296 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 336 390 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 527 527 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 554 640 IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN 672 735 IG-LIKE C2-TYPE DOMAIN 7.
FT DOMAIN 825 1155 PROTEIN KINASE.
FT NP_BIND 831 839 ATP (BY SIMILARITY).
FT BINDING 859 859 ATP (BY SIMILARITY).
FT ACT_SITE 1021 1021 BY SIMILARITY.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 865 865 A -> S (IN REF. 3).
SQ SEQUENCE 1348 AA; 150305 MW; A5E419A476F5DB3 CRC64;

Query Match 25.0%; Score 110.5; DB 1; Length 1348;
Best Local Similarity 35.1%; Pred. No. 6.2e-05;
Matches 27; Conservative 11; Mismatches 26; Indels 13; Gaps 3;

QY 5 CDVSGRPAPYATWQKSH---QRENLMRPDMQGNVVTISIGLVLYNARPEAGLYTCA 61
DB 679 CTVNGVPPPTITWFKNSFTLFEDESGIVLKD---GNKTLT-----IRVRKEDSGLYTLC 728
QY 62 TARNAGLRADFLPLSV 78
DB 729 LACNTLGCKKAEPFSV 745

RESULT 11
FGR1_XENLA STANDARD: PRT: 812 AA.
AC P22182;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR 1 PRECURSOR (EC 2.7.1.112).
GN FGR-1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9104598; PubMed=2172985;
RA Musci T.J., Amaya E., Kirschner M.W.;
"Regulation of the fibroblast growth factor receptor in early Xenopus
```

```
RT embryos.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8365-8369(1990).
CC -1- FUNCTION: RECEPTOR FOR BASIC FIBROBLAST GROWTH FACTOR.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
CC TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U24491; AAA86868.1; -.
DR PIR: A36477; A36477.
DR HSSP: P11362; 1FG1.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR01245; Tyr_kin.
DR Pfam: PF00047; 1g; 3.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SMO0408; IGC2; 3.
DR SMART: SMO0219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW Transferrase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW Signal.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 812 FIBROBLAST GROWTH FACTOR RECEPTOR 1.
FT DOMAIN 21 371 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 372 393 POTENTIAL.
FT DOMAIN 394 812 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 47 107 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 167 233 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 266 344 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 472 761 PROTEIN KINASE.
FT NP_BIND 478 486 ATP (BY SIMILARITY).
FT BINDING 508 508 ATP (BY SIMILARITY).
FT ACT_SITE 617 617 BY SIMILARITY.
FT MOD_RES 648 648 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 812 AA; 90502 MW; B06333BAVEACSC9B CRC64;

Query Match 24.9%; Score 110; DB 1; Length 812;
Best Local Similarity 32.5%; Pred. No. 4e-05;
Matches 26; Conservative 9; Mismatches 39; Indels 6; Gaps 2;

QY 5 CDVSGRPAPYATWQKSH---QRENLMRPDMQGNVVTISIGLVLYNARPEAGLYTCA 63
DB 174 CPANGTPPTLWLK-----NNRPFQODRIGYRVRSQTSWLSLMDSVPSDKGNYTCIV 228
QY 64 RNAAGLRADFLPLSVOREP 83
DB 229 ENKYGAINHTYDLDYVERSP 248
```

RESULT 12
 FGR3_MOUSE STANDARD: PRT; 801 AA.
 AC 061851; 063834;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (EC 2.7.1.112) (FGFR-3)
 GN (HEPARIN-BINDING GROWTH FACTOR RECEPTOR).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RX MEDLINE=92355591; PubMed=1379594;
 RA Ornitz D.M., Leder P.;
 RT "Ligand specificity and heparin dependence of fibroblast growth
 factor receptors 1 and 3."
 RL J. Biol. Chem. 267:16305-16311(1992).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93177694; PubMed=8382556;
 RA Kato O., Hattori Y., Sasaki H., Sakamoto H., Fujimoto K.,
 RA Fujii T., Sugimura T., Terada M.;
 RT "Isolation of the complementary DNA encoding a mouse heparin-binding
 growth factor receptor with the use of a unique kinase insert
 sequence."
 RL Cancer Res. 53:1136-1141(1993).
 RT -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH
 CC FACTORS. PREFERENTIALLY BINDS ACIDIC FGF.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 CC TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: IN EMBRYO, EXPRESSED IN HEART, LUNG, KIDNEY,
 CC SKIN, HEAD AND LIVER BUT NOT IN MUSCLE. IN ADULT, HIGHEST LEVELS
 CC IN BRAIN. ALSO EXPRESSED IN LIVER, LUNG, KIDNEY, TESTIS, OVARY
 CC AND UTERUS. VERY LOW LEVELS IN HEART, THYMUS, SPLEEN AND MUSCLE.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYO FROM MID-GESTATION AND
 CC IN ADULT.
 CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M81342; AAA9535.1; -;
 DR EMBL: S56291; AAB2535.1; -;
 DR HSSP: P11362; IFGI.
 DR MGD: MGI:95524; Fgf3.
 DR InterPro: IPR000719; Euk.pkinase.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_MHC.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00047; Ig_3.
 DR Pfam: PF00069; pkinase.1.
 DR PRINTS: PR00109; TYRKINASE.
 DR SMART: SM00408; IGC2; 3.
 DR SMART: SM00219; TYRC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;

KW Transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 801 FIBROBLAST GROWTH FACTOR RECEPTOR 3.
 FT DOMAIN 21 369 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 370 390 POTENTIAL.
 FT DOMAIN 391 801 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 52 114 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 163 229 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 262 340 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 466 756 PROTEIN KINASE.
 FT NP_BIND 472 481 ATP (BY SIMILARITY).
 FT BINDING 502 502 ATP (BY SIMILARITY).
 FT ACT_SITE 611 611 BY SIMILARITY.
 FT MOD_RES 642 642 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT DISULFID 59 107 POTENTIAL.
 FT DISULFID 170 222 POTENTIAL.
 FT DISULFID 269 333 POTENTIAL.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 684 684 P -> L (IN REF. 2).
 FT CONFLICT 687 687 MISSING (IN REF. 2).
 SQ SEQUENCE 801 AA; 87758 MW; 68BC10212691705 CRC64;

Query Match 24.7%; Score 109; DB 1; Length 801;
 Best Local Similarity 28.2%; Pred. No. 5, 1e-05;
 Matches 24; Conservative 13; Mismatches 32; Indels 16; Gaps 2;

QY 5 CDVSGRPPAVTEKQ-----SHQRNLMRDQNGVNVISGLVYNAPEDAGL 58
 DB 170 CPAGNPTPSISWLKNGKFERGEGIRGLRHQW-----SLWESVPSDRGN 219
 QY 59 YTCARNAGLLRADPPLSVQREP 83
 DB 220 YTCVENKFGSISQTYTLDLVERSP 244

RESULT 13
 FGR3_HUMAN STANDARD: PRT; 806 AA.
 AC P22607; O16294; Q14308;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (EC 2.7.1.112) (FGFR-3).
 GN FGFR3 OR JTK4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX MEDLINE=91142118; PubMed=1847508;
 RA Keegan K., Johnson D.E., Williams L.T., Hayman M.J.;
 RT "Isolation of an additional member of the fibroblast growth factor
 RT receptor family, FGFR-3."
 RL Proc. Natl. Acad. Sci. U.S.A. 88:1095-1099(1991).
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RA Terada M., Shimizu A., Seo M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 76-806 FROM N.A. (LONG ISOFORM), AND TISSUE SPECIFICITY.
 RC TISSUE=Fetal brain;
 RX MEDLINE=92147110; PubMed=1664411;
 RA Thompson L.M., Plummer S., Schalling M., Altherr M.R., Gusella J.F.,
 RA Housman D.E., Wasmuth J.J.;
 RT "A gene encoding a fibroblast growth factor receptor isolated from the

RT Huntington disease gene region of human chromosome 4.";
 RL Genomics 11:1133-1142(1991).
 RN [4]
 RP SEQUENCE OF 614-681 FROM N.A.
 RX MEDLINE-91062389; PubMed-2247464;
 RA Partanen J., Maekelae T.P., Alitalo R., Lehteslahti H.,
 RA Alitalo K.;
 RT "Putative tyrosine kinases expressed in K-562 human leukemia cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8913-8917(1990).
 RN [5]
 RP VARIANT ACH ARG-380.
 RX MEDLINE-94359611; PubMed-8078586;
 RA Rousseau F., Bonneventre J., Legault-Mallet L., Pelet A., Rozet J.-M.,
 RA Martoreaux P., le Merrer M., Munnich A.;
 RT "Mutations in the gene encoding fibroblast growth factor receptor-3
 in achondroplasia.";
 RL Nature 371:252-254(1994).
 RN [6]
 RP VARIANT ACH ARG-380.
 RX MEDLINE-95150025; PubMed-7847369;
 RA Bellus G.A., Hefferon T.W., de Luna R.I., Hecht J.T., Horton W.A.,
 RA Machado M., Kallila I., McIntosh I., Franccomano C.A.;
 RT "Achondroplasia is defined by recurrent G380R mutations of FGFR3.";
 RL Am. J. Hum. Genet. 56:368-373(1995).
 RN [7]
 RP VARIANT ACH CYS-375.
 RX MEDLINE-95278277; PubMed-7758520;
 RA Superti-Furga A., Elch G., Bucher H.U., Wisser J., Gledhill A.,
 RA Gitzelmann R., Steinmann B.;
 RT "A glycine 375-to-cysteine substitution in the transmembrane domain
 of the fibroblast growth factor receptor-3 in a newborn with
 achondroplasia.";
 RL Eur. J. Pediatr. 154:215-219(1995).
 RN [8]
 RP VARIANT TDI CYS-249.
 RX MEDLINE-96154693; PubMed-8589699;
 RA Tavormina P.L., Rimoin D.L., Cohn D.H., Zhu Y.-Z., Shiang R.,
 RA Wasmuth J.J.;
 RT "Another mutation that results in the substitution of an unpaired
 cysteine residue in the extracellular domain of FGFR3 in
 thanatophoric dysplasia type I.";
 RL Hum. Mol. Genet. 4:2175-2177(1995).
 RN [9]
 RP VARIANTS TDI CYS-248 AND CYS-371, AND VARIANT TD2 GLU-650.
 RX MEDLINE-95291326; PubMed-7773297;
 RA Tavormina P.L., Shiang R., Thompson L.M., Zhu Y.-Z., Wilkin D.J.,
 RA Lachman R.S., Wilcox W.R., Rimoin D.L., Cohn D.H., Wasmuth J.J.;
 RT "Thanatophoric dysplasia (types I and II) caused by distinct
 mutations in fibroblast growth factor receptor 3.";
 RL Nat. Genet. 9:321-328(1995).
 RN [10]
 RP VARIANT HYPOCHONDROPLASIA LYS-540.
 RX MEDLINE-95400307; PubMed-7670477;
 RA Bellus G.A., McIntosh I., Smith E.A., Aylsworth A.S., Kallila I.,
 RA Horton W.A., Greenhaw G.A., Hecht J.T., Franccomano C.A.;
 RT "A recurrent mutation in the tyrosine kinase domain of fibroblast
 growth factor receptor 3 causes hypochondroplasia.";
 RL Nat. Genet. 10:357-359(1995).
 RN [11]
 RP VARIANT CROUZON GLU-391.
 RX MEDLINE-96083601; PubMed-7493034;
 RA Meyers G.A., Orlow S.J., Munro I.R., Przybela K.A., Jabs E.W.;
 RT "Fibroblast growth factor receptor 3 (FGFR3) transmembrane mutation
 in Crouzon syndrome with acanthosis nigricans.";
 RL Nat. Genet. 11:462-464(1995).
 RN [12]
 RP CHARACTERIZATION OF VARIANT ACH ARG-380.
 RX MEDLINE-96174812; PubMed-8599935;
 RA Webster M.K., Donoghue D.J.;
 RT "Constitutive activation of fibroblast growth factor receptor 3 by
 the transmembrane domain point mutation found in achondroplasia.";
 RL EMBO J. 15:520-527(1996).
 RN [13]

RP VARIANTS TDI CYS-248; CYS-249; CYS-370 AND CYS-373.
 RX MEDLINE-96254981; PubMed-8845644;
 RA Rousseau F., el Ghouzzi V., Delezoide A.L., Legault-Mallet L.,
 RA le Merrer M., Munnich A., Bonneventre J.;
 RT "Missense FGFR3 mutations create cysteine residues in thanatophoric
 dwarfism type I (TD1).";
 RL Hum. Mol. Genet. 5:509-512(1996).
 RN [14]
 RP VARIANT CRS3 ARG-250.
 RX MEDLINE-971195541; PubMed-9042914;
 RA Muenke M., Gripp K.W., McDonald-McGinn D.M., Gaudenz K.,
 RA Whitaker L.A., Bartlett S.P., Markowitz R.I., Robin N.H., Nwokoro N.,
 RA Mulvihill J.J., Losken H.W., Mulliken J.B., Guttmacher A.E.,
 RA Wilroy R.S., Clarke L.A., Hollway G., Ades L.C., Haan E.A.,
 RA Mulaney D.M., Wall S.A., Wilkie A.O.M., Zackai E.H.;
 RT "A unique point mutation in the fibroblast growth factor receptor 3
 gene (FGFR3) defines a new craniosynostosis syndrome.";
 RL Am. J. Hum. Genet. 60:555-564(1997).
 RN [15]
 RP VARIANT TDI CYS-370.
 RX MEDLINE-99004917; PubMed-9790257;
 RA Katsumata N., Kuno T., Miyazaki S., Mikami S., Nagashima-Miyokawa A.,
 RA Nimura A., Horikawa R., Tanaka T.;
 RT "G370C mutation in the FGFR3 gene in a Japanese patient with
 thanatophoric dysplasia.";
 RL Endocr. J. 45:S171-S174(1998).
 RN [16]
 RP VARIANT HYPOCHONDROPLASIA VAL-538.
 RA Grigelloniane G., Hagenaes L., Ekloef O., Neumeyer L., Haereld P.E.,
 RA Anvret M.;
 RT "A novel missense mutation Ile538Val in the fibroblast growth
 factor receptor 3 in hypochondroplasia.";
 RL Hum. Mutat. 11:333-333(1998).
 RN [17]
 RP VARIANT HYPOCHONDROPLASIA THR-540.
 RX MEDLINE-98112422; PubMed-9452043;
 RA Deutz-Teirlow P.P., Losekoot M., Aalis C.M., Hennemam R.C.M.,
 RA Bakker E.;
 RT "Asn540Thr substitution in the fibroblast growth factor receptor 3
 tyrosine kinase domain causing hypochondroplasia.";
 RL Hum. Mutat. Suppl. 1:S62-S65(1998).
 RN [18]
 RP VARIANT TDI MET-650.
 RA Kitch H., Brodie S.G., Kupke K.G., Lachman R.S., Wilcox W.R.;
 RT "Lys650Met substitution in the tyrosine kinase domain of the
 fibroblast growth factor receptor gene causes thanatophoric dysplasia
 type I.";
 RL Hum. Mutat. 12:362-363(1998).
 CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH
 FACTORS. PREFERENTIALLY BINDS ACIDIC FGF.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, KIDNEY AND TESTIS. VERY
 LOW OR NO EXPRESSION IN SPLEEN, HEART, AND MUSCLE. IN 20-TO 22-
 WEEK OLD FETUSES IT IS EXPRESSED AT HIGH LEVEL IN KIDNEY, LUNG,
 SMALL INTESTINE AND BRAIN, AND TO A LOWER DEGREE IN SPLEEN, LIVER,
 AND MUSCLE.
 CC -1- DISEASE: DEFECTS IN FGFR3 ARE THE CAUSE OF THE AUTOMOSOMAL
 DOMINANT DISEASE ACHONDROPLASIA (ACH); THE MOST FREQUENT FORM OF
 SHORT-LIMB DWAFTISM. ACH IS CHARACTERIZED BY A LONG, NARROW TRUNK,
 SHORT EXTREMITIES, PARTICULARLY IN THE PROXIMAL (RHIZOMELIC)
 SEGMENTS, A LARGE HEAD WITH FRONTAL BOWING, HYPOPLASIA OF THE
 MIDDLE AND A TRIDENT CONFIGURATION OF THE HANDS.
 CC -1- DISEASE: DEFECTS IN FGFR3 ARE A CAUSE OF CROUZON SYNDROME. ALSO
 CALLED CRANIOFACIAL DYOSTOSIS TYPE I (CPDI). CHARACTERIZED BY
 CRANIOSYNOSTOSIS (PREMATURE FUSION OF THE SKULL SUTURES),
 HYPERTELORISM, EXOPHTHALMOS AND EXTERNAL STRABISMOS, PARROT-BEAKED
 NOSE, SHORT UPPER LIP, HYPOPLASTIC MAXILLA, AND A RELATIVE
 MANDIBULAR PROGNATHISM.

-1- DISEASE: DEFECTS IN FGFR3 ARE A CAUSE OF THANATOPHORIC DYSPLASIA (TD) (ALSO KNOWN AS THANATOPHORIC DWAFFISM), THE MOST COMMON NEONATAL LETHAL SKELETAL DYSPLASIA, AFFECTED INDIVIDUALS DISPLAY FEATURES SIMILAR TO THOSE SEEN IN HOMOZYGOUS ACHONDROPLASIA. IT CAUSES SEVERE SHORTENING OF THE LIMBS WITH MACROCEPHALY, NARROW THORAX AND SHORT RIBS. IN THE MOST COMMON SUBTYPE (TD1), FEMUR ARE CURVED, WHILE IN TD2, STRAIGHT FEMURS ARE ASSOCIATED WITH CLOVERLEAF SKULL.

-1- DISEASE: DEFECTS IN FGFR3 ARE A CAUSE OF CRANIOSYNOSTOSIS ADELAIDE TYPE (CRS3), A FORM OF CORONAL SYNOSTOSIS (CS) CHARACTERIZED BY CRANIOSYNOSTOSIS, MIDFACE HYPOPLASIA, DOWNSLANDING PALPEBRAL FISSURES, PROSIS, HIGHLY ARCHED PALATE, MID-TO-MODERATE SENSORINEURAL HEARING LOSS, NORMAL STATURE, BRADYDACTYLY, BROAD BIG TOES. RADIOLOGICALLY, HANDS AND FEET SHOW THIMBLE-LIKE MIDDLE PHALANXES, CONED EPIPHYSES, AND CARPAL AND TARSAL FUSIONS.

-1- DISEASE: DEFECTS IN FGFR3 ARE A CAUSE OF THE AUTOSOMAL DOMINANT DISEASE HYPOCHONDROPLASIA CHARACTERIZED BY DISPROPORTIONATE SHORT STATURE. IT RESEMBLES ACHONDROPLASIA, BUT WITH A LESS SEVERE PHENOTYPE.

-1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR FAMILY.

-1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

Query Match 24.7%; Score 109; DB 1; Length 806;
Best local similarity 28.8%; Pred. No. 5,1e-05;

Matches 23; Conservative 14; Mismatches 37; Indels 6; Gaps 2;

Qy 5 CDVSGRPAPVATWKEQSHQRENLIMRPDMGNVVT-SIGQLVLYNRPEDAGLYTCTA 63
Db 176 CPAAGNPPTPSISWKNRE-----FRGEHRIGIGIKLRHQMWSLYMESVPSDRGNVYCVV 230

Qy 64 RNAAGILRADPPLSVOREP 83
Db 231 ENKRGSIQTYTLVLEKSP 250

RESULT 14

DSCA_HUMAN STANDARD; PRT; 2012 AA.
AC 060469; 060468;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE PRECURSOR (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Plimates; Carnivora; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE-Brain;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Hout Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system."
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Seda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Folley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann N., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kuch J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,

RA Scharie M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ranser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmayer S., Borzys K., Gardiner K., Nizetic D., Francis F.,
RA Leinhardt H., Reinhardt R., Yaspo M.-L.,
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
RN [3]
RP FUNCTION.
RX MEDLINE=20384934; PubMed=10925149;
RA Agatawa K.L., Nakamura S., Tsutsuni Y., Yamakawa K.,
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion."
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM/CHD2-52 (SHOWN HERE)
CC AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC -1- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF023450; AAC17967.1; -
DR EMBL: AF023449; AAC17966.1; -
DR EMBL: AL163283; CAB90464.1; -
DR EMBL: AL163282; CAB90436.1; -
DR EMBL: AL163281; CAB90444.1; -
DR MIM: 602523; -
DR InterPro: IPR001777; FN.III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR Pfam: PF00041; fn3; 6.
DR Pfam: PF00047; Ig; 9.
DR PRINTS: PRO0014; FNTPYELII.
DR SMART: SM00060; FN3; 6.
DR SMART: SM00408; Igc2; 7.
DR SMART: SM00410; Ig_Like; 2.
KW Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW Transmembrane; Alternative splicing.
FT SIGNAL 1 17
FT CHAIN 18 2012
FT DOMAIN 18 1595
FT TRANSMEM 1596 1616
FT DOMAIN 1617 2012
FT DOMAIN 39 109
FT DOMAIN 138 204
FT DOMAIN 239 300
FT DOMAIN 328 392
FT DOMAIN 421 491
FT DOMAIN 518 582
FT DOMAIN 610 676
FT DOMAIN 704 773
FT DOMAIN 802 872
FT DOMAIN 885 972
FT DOMAIN 984 1076
FT DOMAIN 1088 1177
FT DOMAIN 1189 1273
FT DOMAIN 1300 1366
FT DOMAIN 1380 1463
FT DOMAIN 1477 1562
FT DISULFD 46 102
FT DISULFD 145 197

CC	-I- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN TYROSINE PHOSPHATE.
CC	-I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC	-I- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-PROTEIN KINASES.
CC	-----
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).
CC	-----
DR	EMBL; U93306; AAB97508.1; -.
DR	EMBL; U93307; AAB97509.1; -.
DR	HSSP; P06239; 3LCK.
DR	InterPro; IPRO000719; Euk_PKinase.
DR	InterPro; IPRO003006; Ig_MHC.
DR	InterPro; IPRO03598; Ig_C2.
DR	InterPro; IPRO03600; Ig_Like.
DR	InterPro; IPRO01824; Raptor_Tyr_Kin_III.
DR	InterPro; IPRO01245; Tyr_Kin.
DR	Pfam; PF00047; Ig_6.
DR	Pfam; PF00069; pkinase; 2.
DR	SMART; SM00408; IGc2; 1.
DR	SMART; SM00410; IG_Like; 4.
DR	SMART; SM00219; Tyrc; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
DR	PROSITE; PS00240; RECEPTOR_TYR_KIN_DOM; 1.
DR	PROSITE; PS00011; PROTEIN_KINASE_ITK; 1.
KM	Signal; Transferase; Tyrosine-protein kinase; Receptor; Transmembrane; Glycoprotein; Phosphorylation; ATP-binding; Immunoglobulin domain.
FT	SIGNAL 1 19 POTENTIAL.
FT	CHAIN 20 1343 VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2.
FT	DOMAIN 20 760 -EXTRACELLULAR (POTENTIAL).
FT	FT DOMAIN 761 782 CYTOPLASMIC (POTENTIAL).
FT	FT DOMAIN 783 1343 IG-LIKE C2-TYPE DOMAIN 1.
FT	FT DOMAIN 46 109 IG-LIKE C2-TYPE DOMAIN 1.
FT	FT DOMAIN 141 207 IG-LIKE C2-TYPE DOMAIN 2.
FT	FT DOMAIN 239 304 IG-LIKE C2-TYPE DOMAIN 3.
FT	FT DOMAIN 345 400 IG-LIKE C2-TYPE DOMAIN 4.
FT	FT DOMAIN 438 533 IG-LIKE C2-TYPE DOMAIN 5.
FT	FT DOMAIN 560 645 IG-LIKE C2-TYPE DOMAIN 6.
FT	FT DOMAIN 677 740 IG-LIKE C2-TYPE DOMAIN 7.
FT	DOMAIN 830 1158 PROTEIN KINASE.
FT	NP_BIND 836 844 ATP (BY SIMILARITY).
FT	BINDING 844 864 ATP (BY SIMILARITY).
FT	ACT_SITE 1024 1024 BY SIMILARITY.
FT	CARBOHYD 46 46 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 96 96 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 143 143 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 158 158 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 245 245 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 318 318 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 374 374 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 395 395 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 507 507 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 576 576 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 609 609 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 615 615 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 627 627 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 671 671 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 700 700 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD 717 717 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	MOD_RES 1055 1055 PHOSPHORYLATION (AUPO-) (BY SIMILARITY).
SEQ	SEQUENCE 1343 AA; 150393 MW; ADIE509B62D3PF4 CRC64;
Query Match	24.5%; Score 108.5; DB 1; Length 1343;
Best Local Similarity	36.2%; Pred. No. 0.00011;

	Matches	29;	Conservative	9;	Mismatches	33;	Indels	9;	Gaps	3;
QY	5	CDVSGRPPPAVTW	ERKOSHORENLI	MRPDQMYGNVVT	SIGOLVLYNARPE	DAGLYTCTAR	64			
Db	684	CPTSGNPTPLITW	EKDN---ETLV---	EDSGIVLKDGNRN	LTIIRVRKEDG	GLYTCOAC	736			
QY	65	NAAGLLRAD--	FPLSVORE	82						
Db	737	NVLGCAAEFLFI	IEGVQEK	756						

Search completed: February 26, 2002, 01:46:40
Job time: 1184 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:29:35 ; Search time 78.07 seconds
(without alignments)
81.961 million cell updates/sec

Title: US-09-819-136-2_COPY_203_286

Perfect score: 442

Sequence: 1 ASLHCDVSGRPAPVWEKO.....NAAGLIRADPPLSVOREPA 84

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR.68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	128	29.0	898	2	A40114	fasciclin II precursor
2	126.5	28.6	1070	2	JC4593	protein-tyrosine k
3	126	28.5	814	1	A39752	fibroblast growth
4	123.5	27.9	3707	2	S18252	heparan sulfate pr
5	122	27.6	1259	2	S36126	neural cell adhesi
6	122	27.6	1260	1	S05479	neural cell adhesi
7	121.5	27.5	1906	1	S68235	myosin-light-chain
8	121	27.4	1257	1	A41060	neural cell adhesi
9	119	26.9	1344	2	T17346	rig-1 protein - mo
10	117	26.5	483	2	T17346	hypothetical prote
11	116.5	26.4	1612	2	T30805	transmembrane rece
12	115.5	26.1	1651	2	T14160	hypothetical prote
13	115	26.0	2783	2	T34416	hypothetical prote
14	114.5	25.9	4162	2	T26333	connectin/titin -
15	114.5	25.9	5175	2	T20992	hypothetical prote
16	114.5	25.9	5198	2	T43290	hemiscentin precus
17	114.5	25.9	26926	1	I38344	titin, cardiac mus
18	114	25.8	797	2	S38579	fibroblast growth
19	114	25.8	802	1	TVHUF4	fibroblast growth
20	114	25.8	1033	1	S19247	cell adhesion prot
21	112.5	25.5	772	2	T13078	KIAA0992 protein -
22	111.5	25.2	1051	2	A39712	kinase-like protei
23	111.5	25.2	4391	2	A38096	perlecan precursor
24	111	25.1	1091	2	A58532	glial cell membran
25	110.5	25.0	1348	2	S51656	vascular endotheli
26	110	24.9	812	1	A36477	fibroblast growth
27	109	24.7	800	2	TVHUF2	fibroblast growth
28	109	24.7	802	2	A48991	heparin-binding gr
29	109	24.7	801	2	I55363	fibroblast growth

30	109	24.7	806	1	TVHUF3	fibroblast growth
31	109	24.7	1896	2	T08851	Down syndrome cell
32	108.5	24.5	2222	2	T13924	sdh protein - full
33	107	24.2	650	1	JC1450	fibroblast growth
34	107	24.2	729	2	A56795	fibroblast growth
35	107	24.2	822	2	S29840	fibroblast growth
36	107	24.2	871	1	I48696	protein-tyrosine k
37	107	24.2	881	1	I48697	protein-tyrosine k
38	106	24.0	302	2	C36464	fibroblast growth
39	106	24.0	361	2	PN0020	fibroblast growth
40	106	24.0	662	2	C40862	heparin-binding gr
41	106	24.0	725	1	IJMSNG	neural cell adhesi
42	106	24.0	733	2	I49293	fibroblast growth
43	106	24.0	822	1	TVHUF3	fibroblast growth
44	106	24.0	822	1	TVHUF3	fibroblast growth
45	106	24.0	822	2	I49289	fibroblast growth

ALIGNMENTS

```

RESULT 1
A40114
fasciclin II precursor - American bird grasshopper
C:Species: Schistocerca americana (American bird grasshopper)
C>Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 17-Mar-2000
C:Accession: A40114; B31817
R:Harrelson, A.L.; Goodman, C.S.
Science 242, 700-708, 1988
A:Title: Growth cone guidance in insects: fasciclin II is a member of the Immunoglobu
A:Reference number: A40114; MUID:89043938
A:Accession: A40114
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: mRNA
A:Residues: 1-898 <HAR>
R:Snow, P.M.; Zinn, K.; Harrelson, A.L.; McAllister, L.; Schilling, J.; Bastiani, M.J
Proc. Natl. Acad. Sci. U.S.A. 85, 5291-5295, 1988
A:Title: Characterization and cloning of fasciclin I and fasciclin II glycoproteins 1
A:Reference number: A94202; MUID:88276943
A:Accession: B31817
A:Molecule type: mRNA
A:Residues: 423-436 <SNO>
A:Cross-references: GB:J03789; NID:9160848; PID:9160849
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; 1
C:Keywords: transmembrane protein

Query Match          29.0%; Score 128; DB 2; Length 898;
Best Local Similarity 39.0%; Pred. No. 56-06;
Matches 30; Conservative 14; Mismatches 29; Indels 4; Gaps 2;

QY 5 CDVSGRPAPVWEKQSHRENTIMRPDMYGVNVTSIGQLVLYNAREPDAGLYCTAR 64
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 248 CAASGRPVRVYRWIRVDIARD-LTKDGRVSADVL--LGEIRIRVEDANAYSCIAK 303
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 NAAGLIRADPPLSVOR 81
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 304 NAAGTATATVEVTVVVR 320
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
JC4593
protein-tyrosine kinase-related receptor PK7 precursor - human
M:Alternate names: receptor protein tyrosine kinase-like protein (RPTK)
C:Species: Homo sapiens (man)
C>Date: 16-Apr-1996 #sequence_revision 24-May-1996 #text_change 24-Sep-1999
C:Accession: JC4593
R:Park, S.K.; Lee, H.S.; Lee, S.T.
J. Biochem. 119, 235-239, 1996
A:Title: Characterization of the human full-length PK7 cDNA encoding a receptor prot
A:Reference number: JC4593; MUID:97037064
A:Accession: JC4593
A:Molecule type: mRNA

```

A:Residues: 1-1070 <PAR>
 A:Cross-references: GB:U40271; NID:g1322231; PIDN:AA050484.1; PID:g1322232
 C:Comment: This protein is a member of receptor protein tyrosine kinase family, but prot
 C:Genetics:
 A:Gene: GDB:PK7
 A:Cross-references: GDB:134760; OMIM:601890
 A:Map position: 6p21.1-6p12.2
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
 C:Keywords: cell adhesion; extracellular protein; glycoprotein; phosphotransferase; tran
 F:1-30/Domain: signal sequence #status predicted <Sig>
 F:31-1070/Product: protein-tyrosine kinase 7 #status predicted <Mat>
 F:31-703/Domain: extracellular #status predicted <EXC>
 F:704-723/Domain: transmembrane #status predicted <TM>
 F:726-1070/Domain: intracellular #status predicted <INC>
 F:794-1065/Domain: protein kinase homology <KIN>
 F:116,175,184,214,268,283,405,463,567,646/Binding site: carbohydrate (Asn) (covalent) #

Query Match 28.6%; Score 126.5; DB 2; Length 1070;
 Best Local Similarity 30.9%; Pred. No. 8,8e-06;
 Matches 25; Conservative 17; Mismatches 34; Indels 5; Gaps 1;

1 ASLHCDVSGRPAPVTEKOSHORENLIMRPDMYGNVVTSGQLVLYNAREPDAGLYT 60
 DB 609 ALLQCEAGQGPPLQKQKDR-----ILDPKLGPRMHIFQNGSLVHDVAPEDSGRYT 663
 OY 61 CTARNAGLLRADPPLSVVOR 81
 DB 664 CIAGNSCNIKHTEAPLYVVDK 684

RESULT 3
 fibroblast growth factor receptor A1 precursor - African clawed frog
 N:Contains: fibroblast growth factor receptor A1, short splice form; protein-tyrosine ki
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 07-Feb-1992 #sequence_revision 19-Jan-1996 #text_change 16-Jul-1999
 C:Accession: A39752; B39752
 R:Friesele, R.; Dawid, I.B.
 Mol. Cell. Biol. 11, 2481-2488, 1991
 A:Title: cDNA cloning and developmental expression of fibroblast growth factor receptors
 A:Reference number: A39752; MUID:91203867
 A:Accession: A39752
 A:Molecule type: mRNA
 A:Residues: 1-814 <FRI>
 A:Cross-references: GB:M55163; NID:g214893; PIDN:AAA49990.1; PID:g214894
 A:Accession: B39752
 A:Molecule type: mRNA
 A:Residues: 1-30,119-814 <FR2>
 A:Cross-references: GB:M55163
 C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; prote
 C:Keywords: alternative splicing; ATP; autophosphorylation; duplication; glycoprotein; g
 protein kinase
 F:1-21/Domain: signal sequence #status predicted <Sig>
 F:22-814/Product: fibroblast growth factor receptor A1, long splice form #status experim
 F:22-372/Domain: extracellular #status predicted <EXT>
 F:22-30,119-814/Product: fibroblast growth factor receptor A1, short splice form #status
 F:47-102/Domain: immunoglobulin homology <IM1>
 F:125-133/Region: acidic
 F:167-228/Domain: immunoglobulin homology <IM2>
 F:266-339/Domain: immunoglobulin homology <IM3>
 F:373-393/Domain: transmembrane #status predicted <TM>
 F:394-814/Domain: intracellular #status predicted <INT>
 F:472-757/Domain: protein kinase homology <KIN>
 F:480-488/Region: protein kinase ATP-binding motif
 F:514-100,174-226,273-337/Disulfide bonds: #status predicted
 F:76,116,133,177,223,236,260,292,313,326/Binding site: carbohydrate (Asn) (covalent) #st
 F:510,527,619/Active site: Lys, Glu, Asp #status predicted
 F:624,637/Binding site: magnesium (Asn, Asp) #status predicted
 F:650/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicte

Query Match 28.5%; Score 126; DB 1; Length 814;

Best Local Similarity 36.2%; Pred. No. 7.4e-06;
 Matches 29; Conservative 8; Mismatches 37; Indels 6; Gaps 2;

5 CDVSGRPAPVTEKOSHORENLIMRPDMYGNVVTSGQLVLYNAREPDAGLYTCTA 63
 OY 174 CPANGTSPALRFLKLNKE-----FRPDORIGYKVSQSWSLIMDSVPSDKNGNYCTIV 228
 DB 64 RNAAGLLRADPPLSVVOREP 83
 OY 229 ENKIGTINHTYOLDVERSP 248
 DB

RESULT 4
 S18252
 heparan sulfate proteoglycan - mouse
 N:Alternate names: perlecan
 C:Species: Mus musculus (house mouse)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
 C:Accession: S18252; A31917; B31917; S66460
 R:Noonan, D.M.; Fulle, A.; Valente, P.; Cal, S.; Horigan, E.; Sasaki, M.; Yamada, Y.;
 J. Biol. Chem. 265, 22939-22947, 1991
 A:Title: The complete sequence of perlecan, a basement membrane heparan sulfate prote
 adhesion molecule.
 A:Reference number: S18252; MUID:92078153
 A:Accession: S18252
 A:Molecule type: protein
 A:Residues: 1-3707 <NOO>
 A:Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
 R:Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogel, G.; Sasaki, M.; Yamada, Y.; H
 J. Biol. Chem. 263, 16379-16387, 1988
 A:Title: Identification of cDNA clones encoding different domains of the basement mem
 A:Reference number: A92680; MUID:89034110
 A:Accession: A31917
 A:Molecule type: mRNA
 A:Residues: 940-1601 <NO2>
 A:Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
 A:Accession: B31917
 A:Molecule type: mRNA
 A:Residues: 1870-2600 <NO3>
 A:Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
 R:Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A:Title: Structural properties of recombinant domain III-3 of perlecan containing a g
 A:Reference number: S66460; MUID:95377282
 A:Accession: S66460
 A:Molecule type: protein
 A:Residues: 1272-1274, 'X', 1276, 'X', 1278-1279 <SCH>
 C:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G r
 C:Keywords: glycoprotein
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:764-811/Domain: laminin-type EGF-like homology <LEG>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG7>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG7>
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG7>
 F:3163-3198/Domain: EGF homology <EGF>
 F:3270-3423/Domain: laminin G repeat homology <LG2>
 F:3464-3492/Domain: EGF homology <EGF7>
 F:1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predic

Query Match 27.9%; Score 123.5; DB 2; Length 3707;
 Best Local Similarity 39.0%; Pred. No. 7.2e-05;
 Matches 30; Conservative 7; Mismatches 29; Indels 11; Gaps 2;

5 CDVSGRPAPVTEKOSHORENLIMRPDMYGNVVTSGQLVLYNAREPDAGLYTCTAR 64
 DB 2917 CNAAGTPTATITWSKYDGD-----LPPDSRLN-----NMLMLPSVREPEDAGTYCTAT 2965
 OY 65 NAAGLLRADPPLSVVOR 81

A:Accession: A41060
A:Molecule type: mRNA
A:Residues: 1-1257 <HAR>
A:Cross-references: GB:M64296; NID:g186053; PIDN:AA014352.1; PID:g3068548
R:kobayashi, M.; Miura, M.; Asou, H.; Uyemura, K.
Biochim. Biophys. Acta 1090, 238-240, 1991
A:Title: Molecular cloning of cell adhesion molecule LI from human nervous tissue: a
A:Reference number: S18454; MUID:92031698
A:Accession: S18454
A:Molecule type: mRNA
A:Residues: 1-3, 'V', 5-215, 'I', 217-249, 'T', 251-275, 'SV', 278-356, 'E', 358-625, 'V', 627-1257
A:Cross-references: EMBL:X59847; NID:g35009; PIDN:CA44508.1; PID:g35010
A:Note: the authors translated the codon GAA for residue 27 as Gly
R:DiJabail, M.; Mattei, M.G.; Nguyen, C.; Roux, D.; Demangeot, J.; Denizot, F.; Moos, G.
A:Title: The gene encoding LI, a neural adhesion molecule of the immunoglobulin family
A:Reference number: A35331; MUID:90353957
A:Accession: A35331
A:Molecule type: DNA
A:Residues: 332-371 <DJA>
A:Cross-references: GB:M55271
R:Rosenthal, A.; Mackinnon, R.N.; Jones, D.S.C.
Nucleic Acids Res. 19, 5395-5401, 1991
A:Title: PCR walking from microdissection clone M54 identifies three exons from the h
A:Reference number: S21971; MUID:92020233
A:Accession: S21971
A:Molecule type: DNA
A:Residues: 1082-1176 <ROS>
A:Cross-references: EMBL:X58775; NID:g29642; PIDN:CAA41576.1; PID:g29643
A:Accession: S21972
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 353-935, 'V', 937-1176 <RO2>
A:Cross-references: EMBL:X58776; NID:g29644; PIDN:CAB37831.1; PID:g4467833
R:Harper, J.R.; Prince, J.T.; Healy, P.A.; Stuart, J.K.; Nauman, S.J.; Stallcup, W.B.
J. Neurochem. 56, 797-804, 1991
A:Title: Isolation and sequence of partial cDNA clones of human LI: homology of human
A:Reference number: A60223; MUID:91132183
A:Accession: A60223
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1030-1115, 'WLC', 1118-1176, 1181-1257 <HAR>
R:Wolff, J.M.; Frank, R.; Majoo, K.; Spiro, R.C.; Reisfeld, R.A.; Rathjen, F.G.
J. Biol. Chem. 263, 11943-11947, 1988
A:Title: A human brain glycoprotein related to the mouse cell adhesion molecule LI.
A:Reference number: A31072; MUID:88298876
A:Accession: A31072
A:Molecule type: protein
A:Residues: 'Q', 21-36 <WOL>
R:Platzter, M.; Bauer, D.; Drescher, B.
submitted to the EMBL Data Library, March 1995
A:Reference number: H01368
A:Accession: G02506
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1257 <PLA>
C:genetics:
A:Cross-references: EMBL:U052112; NID:g1302657; PIDN:AA051746.1; PID:g1302658
A:Gene: GDB:LLICAM
A:Cross-references: GDB:120133; OMIM:303350; OMIM:308840
A:Map position: Xq28-Xq28
A:introns: 26/1; 31/1; 66/2; 134/1; 175/1; 232/1; 269/2; 331/1; 375/1; 423/1; 460/2;
2/
C:Superfamily: neural cell adhesion molecule LI; fibronectin type III repeat homology
C:Keywords: alternative splicing; cell adhesion; duplication; glycoprotein; transmembr
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1257/Product: neural cell adhesion molecule LI #status predicted <MAT>
F:257-314/Domain: immunoglobulin homology <MM1>
F:532-593/Domain: immunoglobulin homology <MM2>
Query Match 27.4%; Score 121; DB 1; Length 1257;
Best Local Similarity 41.2%; Pired. No 4.2e-05;

Matches 33: Conservative 5; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCDVSGRPPAVTWKQSHQRENLMRPDMYGNVVTISIGOLVLYNARPEADGLYT 60
Db 350 ARLDQOVGRPEPEVTWTRINGIPVEL--AKDKYR---IQRGALILSNQPSDTWTQ 403
OY 61 CTARNAAGLLRADPFLSVYQ 80
Db 404 CEARNRGLLLANAYIVYQ 423

RESULT 9
T14316
t1g-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14316
R:Yuan, S.S.F.; Cox, L.A.; Dasika, G.K.; Lee, E.Y.H.P.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z17975
A:Accession: T14316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1344 <YUA>
A:Cross-references: EMBL:AF060570; NID:g4206385; PID:g4206386; PIDN:AAD11628.1

Query Match 26.9%; Score 119; DB 2; Length 1344;
Best Local Similarity 32.5%; Pred. No. 7.4e-05;
Matches 26; Conservative 15; Mismatches 33; Indels 6; Gaps 2;

OY 2 SLHCDVSGRPPAVTWKQSHQRENLMRPDMYGNVVTISIGOLVLYNARPEADGL 58
Db 343 SFQCTKGNPPAIWQKEGSG---VLLFSSQLQPMGRLLVSPGQLNTEVKIGDGY 399

OY 59 YTCTARNAGLLRADPFLSV 78
Db 400 YVCGAVSVAGSLAKALLEI 419

RESULT 10
T17346
hypothetical protein DKFZp58601624.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17346
R:Diesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18727
A:Accession: T17346
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-483 <DUE>
A:Cross-references: EMBL:AL117666
A:Experimental source: adult uterus; clone DKFZp58601624
C:Genetics:
A:Note: DKFZp58601624.1

Query Match 26.5%; Score 117; DB 2; Length 483;
Best Local Similarity 27.6%; Pred. No. 4e-05;
Matches 24; Conservative 21; Mismatches 24; Indels 18; Gaps 2;

OY 1 ASLHCDVSGRPPAVTWKQSHQRENLMRPDMYGNVVTISIGOLVLYNARPEADGLYT 53
Db 6 ARLECAATGHPNPQIAWQKDGCTDPAARERRMHVMPDDVYF-----FITDVKI 54

OY 54 EDAGLYTCTARNAGLLRADPFLSVYQ 80
Db 55 DDAGVYCTAONSAGSISANATLVLE 81

RESULT 11

T30805
dutt1 protein - mouse
N:Alternate names: transmembrane receptor protein Robol homolog
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30805
R:Wu, M.C.; Lowe, N.; Fordham, R.; Rabbits, P.
submitted to the EMBL Data Library, July 1998
A:Description: The mouse homologue of human DUTT1/H-robol gene: protein sequence and
A:Reference number: Z20879
A:Accession: T30805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1612 <WUM>
A:Cross-references: EMBL:Y17793; NID:e1329712; PID:e1329713; PIDN:CAA76850.1
A:Experimental source: brain
C:Genetics:
A:Gene: dutt1
A:Map position: 16

Query Match 26.4%; Score 116.5; DB 2; Length 1612;
Best Local Similarity 26.9%; Pred. No. 0.00017;
Matches 25; Conservative 14; Mismatches 43; Indels 11; Gaps 1;

OY 2 SLHCDVSGRPPAVTWKQSHQRENLMRPDMYGNVVTISIGOLVLYNARPEADGLYT 61
Db 330 TFQCATGNPQPAIFMRREGSONLLFSYQPPQSSSRFSVSQGTDLITNVQSRSDVGYYIC 389

OY 62 TARNAGLL-----RADPFLSVQREP 83
Db 390 QTLNVAGSITTKAYLEVTVDIADRPPIYRQCP 422

RESULT 12
T14160
transmembrane receptor protein Robol - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14160
R:Kid, T.; Brose, K.; Mitchell, K.J.; Fetter, R.D.; Tessier-Lavigne, M.; Goodman, C.
Cell 92, 205-215, 1998
A:Title: Roundabout controls axon crossing of the CNS midline and defines a novel sub
A:Reference number: Z17897; MUID:98117249
A:Accession: T14160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1651 <KID>
A:Cross-references: EMBL:AF041082; NID:g2811215; PID:g2811216; PIDN:AAC39960.1
C:Function:
A:Description: appears to function as the gatekeeper controlling midline crossing
C:Keywords: transmembrane protein

Query Match 26.1%; Score 115.5; DB 2; Length 1651;
Best Local Similarity 26.9%; Pred. No. 0.00022;
Matches 25; Conservative 14; Mismatches 43; Indels 11; Gaps 1;

OY 2 SLHCDVSGRPPAVTWKQSHQRENLMRPDMYGNVVTISIGOLVLYNARPEADGLYT 61
Db 369 TFQCATGNPQPAIFMRREGSONLLFSYQPPQSSSRFSVSQGTDLITNVQSRSDVGYYIC 428

OY 62 TARNAGLL-----RADPFLSVQREP 83
Db 429 QTLNVAGSITTKAYLEVTVDIADRPPIYRQCP 461

RESULT 13
T34416
hypothetical protein F12F3.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34416

R: Fulton, B.; Wohldmann, P.
 Submitted to the EMBL Data Library, July 1998
 A: Description: The sequence of C. elegans cosmid F12F3.
 A: Reference number: Z21521
 A: Accession: T34416
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-2783 <FUL>
 A: Cross-references: EMBL:U08022; PIDN:AC25886.1; GSPDB:GN00023; CESP:F12F3.2
 A: Experimental source: strain Bristol N2; clone F12F3
 C: Genetics:
 A: Gene: CESP:F12F3.2
 A: Map position: 5
 A: Introns: 45/3; 90/3; 451/3; 509/1; 2313/3; 2341/3; 2378/3; 2414/2; 2453/3; 2474/2; 252

Query Match 26.0%; Score 115; DB 2; Length 2783;
 Best Local Similarity 35.9%; Pred. No. 0.00045;
 Matches 28; Conservative 7; Mismatches 39; Indels 4; Gaps 1;

QY 1 ASLHCDVSGRPAPVATWEEKOSHORENLIIRPDQYGNVVTISIGOLVLYNARPEAGLYTC 60
 |||||
 Db 532 ATLSGDVDGVPSPKQYKDKDELTPSMKYDSFYNE-----GLAELTVKNIVESDACKYT 587
 |||||
 QY 61 CTARNAAGLLRADFPPLSV 78
 |||||
 Db 588 CRAFTNDLGSIMTHAKLSV 605

RESULT 14

T42633
 connectin/titin - chicken (fragment)
 C: Species: Gallus gallus (chicken)
 C: Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C: Accession: T42633
 R: Tajima, H.; Ohtsuka, H.; Kawamura, Y.; Kume, H.; Murayama, T.; Abe, H.; Kimura, S.; Ma
 Biochem. Biophys. Res. Commun. 223, 160-164, 1996
 A: Title: A 11.5-Kb 5'-terminal cDNA sequence of chicken breast muscle connectin/titin re
 A: Reference number: Z22221; MUID:96254045
 A: Accession: T42633
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: mRNA
 A: Residues: 1-4162 <YAJ>
 A: Cross-references: EMBL:D83390; NID:g1513029; PIDN:BA11908.1; PID:g1513030
 A: Experimental source: breast muscle
 C: Keywords: skeletal muscle

Query Match 25.9%; Score 114.5; DB 2; Length 4162;
 Best Local Similarity 32.5%; Pred. No. 0.00079;
 Matches 25; Conservative 14; Mismatches 33; Indels 5; Gaps 1;

QY 2 SLHCDVSGRPAPVATWEEKOSHORENLIIRPDQYGNVVTISIGOLVLYNARPEAGLYTC 61
 |||||
 Db 708 TLECHSGHPQVTVYREYKTES-----SMDFOITFRAGLRLVIREAFADSGRFTC 762

QY 62 TARNNAAGLLRADFPPLSV 78
 |||||
 Db 763 TATNKAGSVSTSCHLV 779

RESULT 15

T20992
 hypothetical protein F15G9.4a - Caenorhabditis elegans
 C: Species: Caenorhabditis elegans
 C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C: Accession: T20992; T24733
 R: Sulston, J.
 Submitted to the EMBL Data Library, December 1994
 A: Reference number: Z19355
 A: Accession: T20992
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA

A: Residues: 1-5175 <WIL>
 A: Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a
 A: Experimental source: clone F15G9
 R: Kershaw, J.
 Submitted to the EMBL Data Library, December 1994
 A: Reference number: Z19929
 A: Accession: T24733
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-5175 <WIL>
 A: Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a
 A: Experimental source: clone T09B9
 C: Genetics:
 A: Gene: CESP:F15G9.4a
 A: Map position: X
 A: Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184
 1; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 303
 1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 50

Query Match 25.9%; Score 114.5; DB 2; Length 5175;
 Best Local Similarity 36.8%; Pred. No. 0.001;
 Matches 28; Conservative 8; Mismatches 25; Indels 15; Gaps 2;

QY 3 LHCDVSGRPAPVATWEEKOSHORENLIIRPDQYGNVVTISIGOLVLYNARPEAGLYTC 62
 |||||
 Db 3236 VHCQVTRGPVYVTKRNGKEIEF-----NPV-----LHTRNATRADEGKYSCT 3280
 |||||
 QY 63 ARNNAAGLLRADFPPLSV 78
 |||||
 Db 3281 ASNEAGTAVADFLIDV 3296

Search completed: February 26, 2002, 01:29:38
 Job time: 417 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:04 ; Search time 69.26 Seconds
(Without alignments)
27.292 Million cell updates/sec

Title: US-09-819-136-2_COPY_203_286

Perfect score: 442
Sequence: 1 ASHCHDSGRPPAVTWKQ.....NAGGLRADPLSVQREPA 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.5	28.4	322	4	US-09-383-586-33
2	122	27.6	92	2	US-08-341-843B-7
3	122	27.6	92	2	US-08-427-497E-12
4	121	27.4	92	2	US-08-341-843B-19
5	121	27.4	92	2	US-08-427-497E-24
6	120.5	27.3	434	4	US-09-540-245A-19
7	117	26.5	1101	3	US-08-986-485-2
8	116.5	26.4	439	4	US-09-383-586-32
9	116.5	26.4	529	4	US-09-383-586-31
10	116.5	26.4	1651	4	US-09-540-245A-18
11	114.5	25.9	1395	4	US-09-540-245A-15
12	113	25.6	868	1	US-08-374-834-1
13	113	25.6	868	2	US-08-644-271-1
14	113	25.6	869	1	US-08-374-834-16
15	113	25.6	869	2	US-08-644-271-29
16	111	25.1	478	5	PCT-US95-08493-15
17	111	25.1	860	5	PCT-US95-08493-19
18	111	25.1	868	5	PCT-US95-08493-21
19	109	24.7	1091	3	US-08-986-485-5
20	109	24.7	801	4	US-09-383-586-30
21	109	24.7	806	4	US-09-383-586-33
22	107.5	24.3	1400	3	US-08-986-485-4
23	106	24.0	300	1	US-07-640-029-5
24	106	24.0	300	4	US-08-439-992A-5
25	106	24.0	302	1	US-07-640-029-6
26	106	24.0	302	1	US-07-921-807B-7
27	106	24.0	302	1	US-07-921-807B-8

ALIGNMENTS

28	106	24.0	302	1	US-08-441-944A-7	Sequence 7, Appl
29	106	24.0	302	1	US-08-441-944A-8	Sequence 8, Appl
30	106	24.0	302	4	US-08-439-992A-6	Sequence 6, Appl
31	106	24.0	351	5	PCT-US93-05703-2	Sequence 2, Appl
32	106	24.0	733	1	US-07-640-029-4	Sequence 4, Appl
33	106	24.0	733	1	US-07-921-807B-6	Sequence 6, Appl
34	106	24.0	733	1	US-08-441-944A-6	Sequence 4, Appl
35	106	24.0	733	1	US-08-439-992A-4	Sequence 4, Appl
36	106	24.0	816	1	US-07-640-029-1	Sequence 1, Appl
37	106	24.0	817	1	US-07-640-029-2	Sequence 1, Appl
38	106	24.0	820	1	US-07-921-807B-3	Sequence 3, Appl
39	106	24.0	820	1	US-08-441-944A-3	Sequence 3, Appl
40	106	24.0	820	1	US-08-166-717D-6	Sequence 6, Appl
41	106	24.0	820	4	US-08-439-992A-1	Sequence 1, Appl
42	106	24.0	822	1	US-07-997-133-1	Sequence 1, Appl
43	106	24.0	822	1	US-07-921-807B-4	Sequence 4, Appl
44	106	24.0	822	1	US-08-459-296-2	Sequence 2, Appl
45	106	24.0	822	1	US-08-441-944A-4	Sequence 4, Appl

```

RESULT 1
US-09-383-586-33
; Sequence 33, Application US/09383586
; Patent No. 6242419
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compounds isolated from stromal cells
; TITLE OF INVENTION: and methods for their use
; FILE REFERENCE: 11000.1037c1
; CURRENT APPLICATION NUMBER: US/09/383,586
; CURRENT FILING DATE: 1999-08-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Human
; US-09-383-586-33

Query Match      28.4% Score 125.5; DB 4; Length 322;
Best Local Similarity 36.7% Pred. No. 7.5e-08;
Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;

QY 3 LHCDSGRPPAVTWKQSHORENLIMRPDQYGVNVVTSIGOLVLYNRPDAGLYTCT 62
DB 168 LKCVASGHRPRITWMKD-----DQALTREAEPR---KKKWTLSIKNLRPDSDSKYTCR 220
QY 63 ARNAGGLRADPLSVQRE 81
DB 221 VSNRAGATNATYKYVDVQIR 239

RESULT 2
US-08-341-843B-7
; Sequence 7, Application US/08341843B
; Patent No. 5872225
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of LIGAM and
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 39

```

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Fay, Sharpe, Beall, Fagan,
3 ADDRESSEE: Minnich & McKee
4 STREET: 1100 Superior Avenue
5 STREET: Suite 700
6 CITY: Cleveland
7 STATE: Ohio
8 COUNTRY: U.S.A.
9 ZIP: 44114-2518
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Diskette, 3.50 inch, 720 KB
12 MEDIUM TYPE: storable
13 COMPUTER: Compaq Prolinea 5100e
14 OPERATING SYSTEM: DOS 5.0
15 SOFTWARE: ASCII
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/341,843B
18 FILING DATE: No. 5872225ember 18, 1994
19 CLASSIFICATION: 435
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: 07/904,991
22 FILING DATE: June 26, 1992
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Minnich, Richard J.
25 REGISTRATION NUMBER: 24,175
26 REFERENCE/DOCKET NUMBER: CWR 2 149-1
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (216) 861-5582
29 TELEFAX: (216) 241-1666
30 TELEX: (216) 980162
31 INFORMATION FOR SEQ ID NO: 7:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 92
34 TYPE: amino acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: amino acids
38 HYPOHETICAL: irrelevant
39 ANTI-SENSE: no
40 ORIGINAL SOURCE:
41 ORGANISM: mouse
42 INDIVIDUAL ISOLATE: 8 day old mouse brain
43 IMMEDIATE SOURCE:
44 LIBRARY: lambda GT 10 and lambda GT11
45 CLONE: synthesis of several clones
46 PUBLICATION INFORMATION:
47 AUTHORS: Moos, M.
48 AUTHORS: Tacke, R.
49 AUTHORS: Scherer, H.
50 AUTHORS: Teplow, D.
51 AUTHORS: Fruh, K.
52 AUTHORS: Schachner, M.
53 TITLE: Neural adhesion molecule L1 is a
54 TITLE: member of the immunoglobulin
55 TITLE: superfamily with binding domains
56 TITLE: similar to fibronectin
57 JOURNAL: NATURE
58 VOLUME: 334
59 ISSUE:
60 PAGES: 701-703
61 DATE: 1988
62 US-08-341-843B-7

Query Match 27.6%; Score 122; DB 2; Length 92;
Best Local Similarity 38.8%; Pred. No. 4.3e-08;
Matches 31; Conservative 7; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCYSGRPAYVWEKQSHORENLIMRPDMYGVVTSIGQLVLYNAREPDAGLYT 60
DB 17 ARLDCCYQGRPOPEITWRINGMSMET--VNKDQKYR---IEGSSILSNVQPTDYMVQ 70
OY 61 CTARNAGLLRADPFLSVQ 80

Db 71 CEARNQHGLLANAVIYVQ 90

1 ||| ||| : |||
2 RESULT 3
3 US-08-427-497E-12
4 Sequence 12, Application US/08427497E
5 Patent No. 5969124
6 GENERAL INFORMATION:
7 APPLICANT: Lemmon, Vance
8 TITLE OF INVENTION: A Method for Characterizing the
9 TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
10 Patent No. 5969124
11 TITLE OF INVENTION: the Nucleotide Sequence
12 NUMBER OF SEQUENCES: 44
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Fay, Sharpe, Beall, Fagan,
15 ADDRESSEE: Minnich & McKee
16 STREET: 1100 Superior Avenue
17 STREET: Suite 700
18 CITY: Cleveland
19 STATE: Ohio
20 COUNTRY: U.S.A.
21 ZIP: 44114-2518
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Diskette, 3.50 inch, 720 KB
24 MEDIUM TYPE: storable
25 COMPUTER: Compaq Prolinea 5100e
26 OPERATING SYSTEM: DOS 5.0
27 SOFTWARE: ASCII
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/427,497E
30 FILING DATE: April 24, 1995
31 CLASSIFICATION: 435
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: 07/904,991
34 FILING DATE: June 26, 1992
35 ATTORNEY/AGENT INFORMATION:
36 NAME: Minnich, Richard J.
37 REGISTRATION NUMBER: 24,175
38 REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: (216) 861-5582
41 TELEFAX: (216) 241-1666
42 TELEX: (216) 980162
43 INFORMATION FOR SEQ ID NO: 12:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 92
46 TYPE: amino acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49 MOLECULE TYPE: amino acids
50 HYPOHETICAL: irrelevant
51 ANTI-SENSE: no
52 ORIGINAL SOURCE:
53 ORGANISM: mouse
54 INDIVIDUAL ISOLATE: 8 day old mouse brain
55 IMMEDIATE SOURCE:
56 LIBRARY: lambda GT 10 and lambda GT11
57 CLONE: synthesis of several clones
58 PUBLICATION INFORMATION:
59 AUTHORS: Moos, M.
60 AUTHORS: Tacke, R.
61 AUTHORS: Scherer, H.
62 AUTHORS: Teplow, D.
63 AUTHORS: Fruh, K.
64 AUTHORS: Schachner, M.
65 TITLE: Neural adhesion molecule L1 is a
66 TITLE: member of the immunoglobulin
67 TITLE: superfamily with binding domains
68 TITLE: similar to fibronectin
69 JOURNAL: NATURE

VOLUME: 334
ISSUE:
PAGES: 701-703
DATE: 1988
US-08-427-497E-12

Query Match 27.6%; Score 122; DB 2; Length 92;
Best Local Similarity 38.8%; Pred. No. 4.3e-08;
Matches 31; Conservative 7; Mismatches 36; Indels 6; Gaps 2;

QY 1 ASLHCDVSGRPPAVTWKQSHORENLIMRPOMYGNVVTSGOLVLYNARPEAGLYT 60
DB 17 ARLDQVQGRPPQPEVWTRNGIPVEEL--AKQDKYR-----IQGALILSNVQPSDIMVYQ 70
QY 61 CTARNAGLLRADFPLSVYQ 80
DB 71 CEARNHGILLANAYIYVYQ 90

RESULT 4
US-08-341-843B-19

; Sequence 19, Application US/08341843B
; Patent No. 5872225
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5872225
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; MEDIUM TYPE: storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/341,843B
; FILING DATE: No. 5872225ember 18, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/904,991
; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acids
; HYPOTHEICAL: irrelevant
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens

INDIVIDUAL ISOLATE: 17-18 week fetus
IMMEDIATE SOURCE:
LIBRARY: Stratagene cDNA Library 936206
CLONE: synthesis of 4 clones
PUBLICATION INFORMATION:
AUTHORS: Hlavín, Mary Louise
AUTHORS: Lemmon, Vance

TITLE: Molecular structure and functional
TITLE: testing of human L1CAM: an interspecies
TITLE: comparison
JOURNAL: GENOMICS
VOLUME: 11
ISSUE:
PAGES: 416-423
DATE: 1991
US-08-341-843B-19

Query Match 27.4%; Score 121; DB 2; Length 92;
Best Local Similarity 41.2%; Pred. No. 5.7e-08;
Matches 33; Conservative 5; Mismatches 36; Indels 6; Gaps 2;

QY 1 ASLHCDVSGRPPAVTWKQSHORENLIMRPOMYGNVVTSGOLVLYNARPEAGLYT 60
DB 17 ARLDQVQGRPPQPEVWTRNGIPVEEL--AKQDKYR-----IQGALILSNVQPSDIMVYQ 70
QY 61 CTARNAGLLRADFPLSVYQ 80
DB 71 CEARNHGILLANAYIYVYQ 90

RESULT 5
US-08-427-497E-24

; Sequence 24, Application US/08427497E
; Patent No. 5969124
; GENERAL INFORMATION:
; APPLICANT: Lemmon, Vance
; TITLE OF INVENTION: A Method for Characterizing the
; TITLE OF INVENTION: Nucleotide Sequence of L1CAM and
; Patent No. 5969124
; TITLE OF INVENTION: the Nucleotide Sequence
; TITLE OF INVENTION: Characterized Thereby
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; MEDIUM TYPE: storable
; COMPUTER: Compaq Prolinea 5100e
; OPERATING SYSTEM: DOS 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/427,497E
; FILING DATE: April 24, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/904,991
; FILING DATE: June 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: CWR 2 149-3-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162

INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 92
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acids
HYPOTHEICAL: irrelevant
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
INDIVIDUAL ISOLATE: 17-18 week fetus
IMMEDIATE SOURCE:
LIBRARY: Stratagene cDNA library 936206
CLONE: synthesis of 4 clones
PUBLICATION INFORMATION:
AUTHORS: Hlavin, Mary Louise
AUTHORS: Lemmon, Vance
TITLE: Molecular structure and functional
TITLE: testing of human hICAM: an interspecies
TITLE: comparison.
JOURNAL: GENOMICS
VOLUME: 11
ISSUE: 416-423
PAGES: 416-423
DATE: 1991
US-08-427-497E-24

Query Match 27.4%; Score 121; DB 2; Length 92;
Best Local Similarity 41.2%; Pred. No. 5.7e-08;
Matches 33; Conservative 5; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCDSVSGRPPAVTWKSHQRENLMRPDMYGNVVTSGQLVLYNARPPDGLY 60
DB 17 ARDCOVQGRPOEVTRIRINGIPVEEL--AKDQYR-----IQRGALISVQPSDVIWVQ 70
OY 61 CTARNAGLLRADPFLSVQ 80
DB 71 CEARNRHGILLANAYIVVQ 90

RESULT 6
US-09-540-245A-19
Sequence 19, Application US/09540245A
Patent No. 6270984
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/540,245A
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 60/065,544
PRIOR FILING DATE: 1997-11-14
PRIOR APPLICATION NUMBER: 60/081,057
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 434
TYPE: PRT
ORGANISM: human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (285)..(396)
OTHER INFORMATION: note="Xaa signifies gap in sequence"
US-09-540-245A-19

Query Match 27.3%; Score 120.5; DB 4; Length 434;

Best Local Similarity 32.3%; Pred. No. 4.6e-07;
Matches 30; Conservative 12; Mismatches 36; Indels 15; Gaps 3;
OY 5 CDVSGRPPAVTWKSHQRENLM--RPDMYGNVVTSGQLVLYNARPPDGLYTC 62
DB 13 CETKGNPQAPVAFWQKESGQ--NLLFPNQPOQPSKCSVPTGLTITNIGRSDAGYICQ 70
OY 63 ARNAGLLRA-----DFPLSVQREPA 84
DB 71 ALTVAGSILAKALEVTDVLTDRPPITLIGPA 103

RESULT 7
US-08-986-485-2
Sequence 2, Application US/08986485
Patent No. 6046030
GENERAL INFORMATION:
APPLICANT: WU, SHUTIAN
APPLICANT: SWEET, RAYMOND
APPLICANT: TRUNEH, ALEMSEGED
TITLE OF INVENTION: A HUMAN LIG-1 HOMOLOG (HLIG-1)
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,485
FILING DATE: 08-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/059,448
FILING DATE: 22-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1101 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-986-485-2

Query Match 26.5%; Score 117; DB 3; Length 1101;
Best Local Similarity 27.6%; Pred. No. 4e-06;
Matches 24; Conservative 21; Mismatches 24; Indels 18; Gaps 2;

OY 1 ASLHCDSVSGRPPAVTWKSHQRENLMRPDMYGNVVTSGQLVLYNARPP 53
DB 624 ARLECAATGHPNPQIAMQDGTDPFAERKRNHYVPPDDVF-----FTIDVKI 672
OY 54 EDAGLYTCTARNAGLLRADPFLSVQ 80
DB 673 DDAGVYSTAONSAGSISANATLVILE 699

RESULT 8

```
US-09-383-586-32
; Sequence 32, Application US/09383586
; Patent No. 6242419
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevlin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compounds isolated from stromal cells
; FILE REFERENCE: 11000.1037c1
; CURRENT APPLICATION NUMBER: US/09/383,586
; CURRENT FILING DATE: 1999-08-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Mouse
US-09-383-586-32

Query Match          26.4%; Score 116.5; DB 4; Length 439;
Best Local Similarity 32.5%; Pred. No. 1.5e-06;
Matches 27; Conservative 12; Mismatches 29; Indels 15; Gaps 2;

QY 3 LHCDVSGRPPAVTWKSHORENLIMRPDMYGNVVTISIGQ---LVLYNARPEADAGL 58
DB 166 LKCVASGHRPDIIMW-----MKDDQTLTHLEASEHRRKKWTLISLKNLKPEDSGK 214
QY 59 YTCARNAGLLRADPPLSVQR 81
DB 215 YTCRVSNKAGAINATYKVDIQR 237

US-09-540-245A-18
; Sequence 18, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1651
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-18

Query Match          26.4%; Score 116.5; DB 4; Length 1651;
Best Local Similarity 26.9%; Pred. No. 7.7e-06;
Matches 25; Conservative 14; Mismatches 43; Indels 11; Gaps 1;

QY 2 SLHCDVSGRPPAVTWKSHORENLIMRPDMYGNVVTISIGQLVLYNARPEADAGLYTC 61
DB 369 TFOCEATGNPOPAIFWRREGSONLFSYOPSSRSFVSQGTDLITTVQSRDVGYYIC 428
QY 62 TARNAGLL-----RADPPLSVQREP 83
DB 429 QTLNAGSIITTKAYLEVTVIADRPPIVIRQGP 461

RESULT 11
US-09-540-245A-15
; Sequence 15, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-245A-15

Query Match          25.9%; Score 114.5; DB 4; Length 1395;
Best Local Similarity 32.5%; Pred. No. 1.1e-05;
Matches 27; Conservative 14; Mismatches 37; Indels 5; Gaps 2;

QY 3 LHCDVSGRPPAVTWKSHORENLIMRPDMYGNVVTISIGQLVLYNARPEADAGLYTC 62
DB 166 LKCVASGHRPDIIMW-----MKDDQTLTHLEASEHRRKKWTLISLKNLKPEDSGK 214
QY 59 YTCARNAGLLRADPPLSVQR 81
DB 215 YTCRVSNKAGAINATYKVDIQR 237
```

Db 364 LPCMASGNPPSPVFWTKRG---VSTLMFPNSSHGROYVAAADGTQITDPRQDEGGYVCS 420
OY 63 ARNA--AGLLRADPLSYQREP 83
Db 421 AFSVVDSSSTVRVFLQVSSYDERP 443

RESULT 12
US-08-374-834-1
; Sequence 1, Application US/08374834
; Patent No. 5656473
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,834
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,658
; FILING DATE: 21-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 1904
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 345-7400
; TELEFAX: (914) 345-7721
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 868 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-374-834-1

Query Match 25.6%; Score 113; DB 1; Length 868;
Best Local Similarity 37.7%; Pred. No. 9.4e-06;
Matches 26; Conservative 8; Mismatches 25; Indels 10; Gaps 2;
OY 1 ASLHCDVSGRPAPVATWEK-QSHQRENILMRPDMGNVVTISGOLVLYNAREPDAGLY 59
Db 138 AVLPCITMGPKPSVSWIKGDSALREN-----SRIVLESGSLRIHNVOKEKEDAGY 188
OY 60 TCTARNAG 68
Db 189 RCYAKNSLG 197
RESULT 13
US-08-644-271-1
; Sequence 1, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US8N 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 868 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-644-271-1

Query Match 25.6%; Score 113; DB 2; Length 868;
Best Local Similarity 37.7%; Pred. No. 9.4e-06;
Matches 26; Conservative 8; Mismatches 25; Indels 10; Gaps 2;
OY 1 ASLHCDVSGRPAPVATWEK-QSHQRENILMRPDMGNVVTISGOLVLYNAREPDAGLY 59
Db 138 AVLPCITMGPKPSVSWIKGDSALREN-----SRIVLESGSLRIHNVOKEKEDAGY 188
OY 60 TCTARNAG 68
Db 189 RCYAKNSLG 197

RESULT 14
US-08-374-834-16
; Sequence 16, Application US/08374834
; Patent No. 5656473
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,834
; FILING DATE: 19-JAN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/095,658
FILING DATE: 21-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 190A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 345-7400
TELEFAX: (914) 345-7721
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 869 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-374-834-16

Query Match 25.6%; Score 113; DB 1; Length 869;
Best Local Similarity 37.7%; Pred. No. 9,4e-06;
Matches 26; Conservative 8; Mismatches 25; Indels 10; Gaps 2;

OY 1 ASLHCDVSGRPPPAYTWK-QSHORENLIRPDQMGVNVVTSIGQLVYNARPEAGLY 59
DB 138 AVLPTCTMGNPKPSVSWINGDSPLREN-----SRTAVLESGSLRIHNQKEDAGQY 188
OY 60 TCTARNAAG 68
DB 189 RCVAKNSLG 197

RESULT 15
US-08-644-271-29
Sequence 29, Application US/08644271
Patent No. 5814478
GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
TITLE OF INVENTION: AND LIGANDS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 869 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-644-271-29

Query Match 25.6%; Score 113; DB 2; Length 869;
Best Local Similarity 37.7%; Pred. No. 9,4e-06;
Matches 26; Conservative 8; Mismatches 25; Indels 10; Gaps 2;

OY 1 ASLHCDVSGRPPPAYTWK-QSHORENLIRPDQMGVNVVTSIGQLVYNARPEAGLY 59
DB 138 AVLPTCTMGNPKPSVSWINGDSPLREN-----SRTAVLESGSLRIHNQKEDAGQY 188
OY 60 TCTARNAAG 68
DB 189 RCVAKNSLG 197

Search completed: February 26, 2002, 01:28:06
Job time: 360 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:42 ; Search time 162.6 Seconds
(without alignments)
38.267 Million cell updates/sec

Title: US-09-819-136-2_COPY_203_286
Perfect score: 442
Sequence: 1 ASLHCVDSRPPPAVTVWEKQ.....NAAGLRADFPPLSVQREPA 84

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_1101: *
2: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	126.5	28.6	1070	AAW08747	Human colon carcin
2	125.5	28.4	322	AA19116	Homologue of polyp
3	125.5	28.4	365	AAW2511	Human protein sequ
4	125.5	28.4	371	AAW0429	Human polypeptide
5	125.5	28.4	504	AAW24066	Human fibroblast
6	125.5	28.4	504	AAW2864	Human fibroblast
7	125.5	28.4	504	AAW6656	Membrane-bound pro
8	125.5	28.4	504	AAW65179	Human PRO943 (UNO4
9	125.5	28.4	608	AAW6264	Human MANGO 003 SE
10	125.5	28.4	608	AAW38643	Human polypeptide
11	121	27.4	545	AAW62783	Amino acid sequenc

12	121	27.4	579	22	AAW62782	Amino acid sequenc
13	121	27.4	1257	20	AAW74152	Human L1 cell adhe
14	120.5	27.3	434	20	AAW13567	Human Robo 2 polyp
15	120.5	27.3	434	20	AAW08405	Human partial ROBO
16	117.5	26.6	932	22	AAE05252	Mouse Nope (neighb
17	117.5	26.6	1252	22	AAE05251	Mouse Nope (neighb
18	117	26.5	767	22	AAW25696	Human protein sequ
19	117	26.5	1101	21	AAW08008	Human HLTG-1 prote
20	116.5	26.4	101	21	AAW57131	Human prostate can
21	116.5	26.4	439	21	AAW19115	Polypeptide isolat
22	116.5	26.4	529	21	AAW19114	Human T85 protein.
23	116.5	26.4	753	20	AAW83927	Human ROBO1 protei
24	116.5	26.4	1649	20	AAW08404	Human Robo1 protei
25	116.5	26.4	1651	20	AAW13566	Human Robo 1 polyp
26	115	26.0	1853	21	AAW3668	Protein 608 sequen
27	115	26.0	2387	21	AAW3665	Mechanical stress
28	115	26.0	2597	21	AAW3664	Mechanical stress
29	114.5	25.9	1395	20	AAW13563	Drosophila Robo 1
30	114.5	25.9	1395	20	AAW08401	Drosophila sp. ROB
31	114.5	25.9	3117	21	AAW5667	Sequence g1/332818
32	114.5	25.9	26926	22	AAW05396	Human titin (conn
33	114	25.8	1059	20	AAW08095	Human PRO335 prote
34	114	25.8	1059	20	AAW13393	Amino acid sequenc
35	114	25.8	1059	21	AAW70672	Human PRO335 prote
36	114	25.8	1059	22	AAW00825	Human immune respo
37	114	25.8	1059	22	AAW80261	Human PRO335 prote
38	114	25.8	1119	20	AAW08114	Human PRO326 prote
39	114	25.8	1119	21	AAW70674	Human PRO326 prote
40	114	25.8	1119	22	AAW12347	Human PRO326 polyp
41	114	25.8	1119	22	AAW00827	Human immune respo
42	114	25.8	1119	22	AAW80263	Human PRO326 prote
43	114	25.8	1119	22	AAW81652	Human PRO326 polyp
44	114	25.8	1120	20	AAW13395	Amino acid sequenc
45	114	25.8	1744	22	AAW39040	Human polypeptide

ALIGNMENTS

RESULT 1	AAW08747	standard; Protein: 1070 AA.
ID	AAW08747	
XX	AAW08747	
AC	AAW08747	
XX	18-MAR-1997	(first entry)
DT	18-MAR-1997	
XX		
DE	Human colon carcinoma kinase 4 (CCK-4).	
XX		
KW	Colon carcinoma kinase 4; CCK-4; receptor tyrosine kinase;	
KW	Signal transduction; colon cancer; diagnosis; therapy.	
OS		
XX	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Peptide	1..26
FT	Protein	/label= Stg_peptide
FT		27..1070
FT	Domain	/label= Mat_protein
FT		27..704
FT	Region	/label= Extracellular_domain
FT		53..101
FT		/label= Ig_domain_loop
FT		/note= "immunoglobulin domain loop bounded by Cys residues"
FT	Region	150..200
FT		/label= Ig_domain_loop
FT		/note= "immunoglobulin domain loop bounded by Cys residues"
FT	Region	246..301
FT		/note= "immunoglobulin domain loop bounded by Cys residues"
FT	Region	343..391

CC from arthritis, inflammatory bowel disease and cardiac failure and a
CC fibroblast growth factor-mediated disorder. It is also useful in assays
CC to determine biological activity, to raise antibodies, to isolate
CC corresponding ligands or receptors, to quantify levels of protein or
CC cognate corresponding ligand or receptors, as antiinflammatory agents,
CC and in compositions for the treatment of skin, connective tissue and
CC immune system diseases. The polynucleotide is useful as marker for
CC tissue, as a chromosome marker or tags in the identification of a
CC genetic disorder.

Query Match	28.4%	Pred. 135.5	DB 21	Length 322
Best Local Similarity	36.7%	Pred. No. 6.2e-07		
Matches 29	Conservative	9	Mismatches 34	Indels 7
				Gaps 2

OY		3	LHCVSGGPPAPVAMKESHSQREILIRPDQMGVVNTYSIGOLIVYNARPEADGTYCT	62
Dd		168	Ikcvaashnpdrptclcmkd----dqalcrrpeaepr---kkwllslknlrlpredsgkyctr	220
OY		63	ARNNAGLLRADPPLSVVOR	81
	:		:	
Dd		221	vsnrfagaaalnalykvdvilqr	239
	:		:	

RESULT	3
AAM25511	
ID	AAM25511 standard; Protein; 365 AA

AC	AAM25511;
XX	
DT	16-OCT-2001 (first entry)

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antithyroid; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; viricide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antineoplastic; anemia;
KW antilegionella; haemostatic; vulnery; antifungal; osteopathic; ezema;
KW dermatological; antiallergic; antisthmatic; antidiabetic; cystostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antidiaplytic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathologic; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

PT Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 20; Page 210; 1217pp; English.
PS

CC AAH99161 to AAH99904 encode the human proteins given in AAH5225 to
CC AAH59563. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antihumanistic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virocidic; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antihaemic; antipregnant; haemostatic; vulnereary;
CC antitumor; osteoprotic; dermatologic; antiallergic; antiaesthetic;
CC antibacterial; cytostatic; neuroprotective; antidepressant; nocrotropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmune, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurologic disorders.

Query Match	28.48;	Score 125.5;	DB 22;	Length 365;
Best Local Similarly	36.78;	Pred. No. 7.2e-07;		
Matches 29; Conservative	9;	Mismatches 34;	Indels 7;	Gaps 2;

OY 3 LHCDSVGRPEATWYKSHORENIIMPRDOMYGNNVTSTIGOLIVYNAPBEDAGLYTCT 62
| | | | : : : : | | | |
Db 43 lkcvaaghpdpdtwmkd-----dgalttpeaaep---kkwtlslknlrpedsgkyctcr 95

OY 63 ARNNAGLLRADPEPLSVYOR 81
| | : : : : : :
Db 96 vsnrnagalnatkyvdvlqr 114

RESULT	4
AAM40429	
ID	AAM40429 standard; Protein; 371 AA

AC	AAM40429;
XX	
DT	22-OCT-2001 (first entry)

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokine; chemokine; drug screening; arthritis; inflammation;
KW leukaemia.

```

XX PF      06-JAN-2000; 2000MO-US00376.
XX XX
PR      08-MAR-1999;          99MO-US05028.
PR      02-JUN-1999;          99MO-US12252.
PR      23-JUN-1999;          99US-0141037.
PR      07-JUL-1999;          99US-0143048.
PR      26-JUL-1999;          99US-0145698.
PR      30-NOV-1999;          99MO-US28313.
PR      20-DEC-1999;          99MO-US30911.
PR      05-JAN-2000;          2000MO-US00219.
XX XX
PA      (GETH ) GENENTECH INC.
PI      Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA.
PI      Watanabe CK, Wood WT.
XX XX      WPI: 2000-572270/53.
DR      N-PSDB; AAC583376.
XX XX
PT      Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT      treatment, diagnosis and prevention of cancer -
XX XX
PS      Claim 61; Fig 20; 286pp; English.
XX XX
CC      The present invention describes an isolated antibody that binds to
CC      one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC      PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC      PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC      PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC      PRO2145 OR PRO2188. PRO antagonists can be used to inhibit tumour cell
CC      growth. The PRO polypeptides and nucleotides are useful in the
CC      treatment, diagnosis and prevention of cancer. The antibodies and other
CC      anti-tumour compounds maybe used to treat various conditions, including
CC      those characterised by overexpression and/or activation of the amplified
CC      PRO genes. Exemplary conditions or disorders to be treated with such
CC      antibodies and other compounds include benign or malignant tumours
CC      (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC      colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC      carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC      leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC      glial, astrocytal, hypochthalmic and other glandular, macrophagal,
CC      epithelial, stromal and blastocellic disorders, and inflammatory,
CC      angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC      primers and hybridisation probes used in the isolation of the human PRO
CC      sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC      PRO polynucleotide and protein sequences given in the exemplification of
CC      the present invention.
XX XX
SQ      Sequence      504 AA;
XX XX
Query Match              28.4%; Score 125.5; DB 21; Length 504;
Best Local Similarity    36.7%; Pred. No. 1.le=06;
Matches   29; Conservative     9; Mismatches 34; Indels   7; Gaps   2
OY      3 LHCDSGGPPRAVTVTEKSHORENIMRPDQIGVVNTSICGLVLYNARPDAAGTCT 62
        | | | | | | | | | | : | | : | | | | | | | | | | | | | | | |
DB      170 Ikcvasghnpdrpltkmkd---dgaltirpeaaep-----kkkwlslsklnrpedsglytor 222
        | | | | | | | | | | : | | : | | | | | | | | | | | | | | | |
OY      63 ARNAAGLRADPDLISVYOR 81
        | | | | | | | | | | : | | : | | | | | | | | | | | | | | | |
DB      223 vsnragainatykvavldgr 241
        | | | | | | | | | | : | | : | | | | | | | | | | | | | | | |

RESULT      6
AAAY92864
ID      AAAY92864 standard; Protein: 504 AA.
AC      AAAY92864;
XX XX
DT      29-AUG-2000 (first entry)
XX XX

```

DE Human fibroblast growth factor receptor 5.
XX
KW FGFR-5: fibroblast growth factor receptor 5; cytosolic; anti-sclerotic;
KW immunomodulatory; gastrointestinal; virucide; anti-inflammatory;
KW anti-ischemic; anti-atherosclerosis; angiogenic; endocrine;
KW anti-diabetic; gene therapy.
XX
OS Homo sapiens.
XX
FH Key
FT Peptide
FT /label- leader_sequence
FT 23..37
FT /label- antigenic
FT 25..504
FT Protein
FT /label- mature_protein
FT 25..117
FT /label- extracellular
FT /note- "immunoglobulin domain I"
FT 39..48
FT /label- antigenic
FT 51..59
FT /label- antigenic
FT 62..76
FT /label- antigenic
FT 81..97
FT /label- antigenic
FT 101..104
FT /label- antigenic
FT 118..143
FT /label- acid_box_domain
FT 119..170
FT /label- antigenic
FT 144..239
FT /label- extracellular
FT /note- "immunoglobulin domain II"
FT 176..204
FT /label- antigenic
FT 209..228
FT /label- antigenic
FT 238..247
FT /label- antigenic
FT 240..357
FT /label- extracellular
FT /note- "immunoglobulin domain III"
FT 259..262
FT /label- antigenic
FT 268..275
FT /label- antigenic
FT 282..302
FT /label- antigenic
FT 307..320
FT /label- antigenic
FT 326..334
FT /label- antigenic
FT 356..375
FT /label- antigenic
FT 358..373
FT /label- membrane_proximal_domain
FT 374..403
FT /label- transmembrane_domain
FT 401..435
FT /label- antigenic
FT 404..504
FT /label- intracellular_domain
FT 440..443
FT /label- antigenic
FT 446..455
FT /label- antigenic
FT 462..475
FT /label- antigenic
FT 483..496
FT /label- antigenic

XX
PN MO200024756-A1.
XX
PD 04-MAY-2000.
XX
PF 17-JUN-1999; 99WO-US13620.
XX
PR 23-OCT-1998; 98US-0105465.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Young PE;
XX
DR WPI: 2000-387035/33.
DR N-PSDB: AAA28842.
XX
PT Nucleic acids encoding fibroblast growth factor-5 useful for the
PT prevention, diagnosis and treatment of conditions associated with
PT tissue repair and aberrant cell functions, e.g. cell survival and
PT proliferation
XX
PS Claim 11; Fig 1A-C; 182pp; English.
XX
CC This is the fibroblast growth factor receptor protein, FGFR-5. The
CC FGFR-5 protein and DNA may be used in the prevention, treatment and
CC diagnosis of diseases or conditions associated with inappropriate FGFR-5
CC expression and activity. For example, the nucleic acids (and vectors
CC containing them) and the FGFR-5 polypeptide may be used to treat
CC disorders associated with increased or decreased cell survival (such as
CC cancers (e.g. leukemia, colonic cancer, testicular cancer and follicular
CC lymphomas), autoimmune disorders (e.g. multiple sclerosis and Crohn's
CC disease) viral infections (e.g. herpes viruses), inflammation, graft
CC versus host disease, acute and chronic graft rejection, ischemic injuries
CC and atherosclerosis), activation, secretion, migration, differentiation
CC and proliferation, diseases associated with defects in wound healing,
CC mucostosis, defects of angiogenesis, immune dysfunction, endocrine
CC dysfunction and insulin secretion disorders. Anti-FGFR-5 antibodies may
CC also be used as diagnostic agents for detecting the presence of FGFR-5
CC polypeptides in samples.
XX
SQ Sequence 504 AA:

Query Match 28.4%; Score 125.5; DB 21; Length 504;
Best Local Similarity 36.7%; Pred. No. 1,1e-06;
Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;

QY 3 LHCDSGRPPAVTWKSHORENLTMRPDQYGNVVTISIGLVLYNRPDAGLYTCT 62
DB 170 Ikcvasghprpdltwmkd----dgaltrpeaapr---kkwtlslknlrpedsqkytcr 222
QY 63 ARNAAGLLRADPPLSVQR 81
DB 223 vsnragalnatykvdyidr 241

RESULT 7
AAV66656
ID AAV66656 standard; protein; 504 AA.
XX
AC AAV66656;
XX
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO943.
XX
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS Homo sapiens.
XX
PN WO963088-A2.
XX

PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99WO-US12252.
XX
PR 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 05-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089807.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.

PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 07-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

(GEETH) GENENTECH INC.
XX
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WL, Yuan J;
XX

DR WPI: 2000-072883/06.
 DR N-PSDB; AAZ64984.
 XX
 PT Membrane-bound proteins and related nucleotide sequences -
 XX
 PS claim 12; Fig 70; 822pp; English.
 XX
 CC The invention provides membrane-bound PRO polypeptides and
 CC polynucleotides encoding them. The PRO sequences of the invention were
 CC identified based on extracellular domain homology screening. The PRO
 CC sequences have homology with proteins including LDL receptors, TIE
 CC ligands and various enzymes. The membrane-bound proteins and receptor
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
 CC immunoadhesins, for instance, can be used as therapeutic agents to block
 CC receptor-ligand interactions. The membrane-bound proteins can also be
 CC employed for screening of potential peptide or small molecule inhibitors
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences
 CC are useful as hybridization probes, in chromosome and gene mapping and in
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
 CC will also be useful for the preparation of PRO polypeptides, especially
 CC by recombinant techniques.
 CC
 XX
 SO Sequence 504 AA:
 Query Match 28.4%; Score 125.5; DB 21; Length 504;
 Best Local Similarity 36.7%; Pred. No. 1.1e-06;
 Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;
 QY 3 LHCDVSGRPPAVTWKQSHORENLIMRPPQKGNVVTISIGOLVLYNARPDAGLYTCT 62
 DB 170 lkcvasghprpdltwmkd----dgaltrpeaaep---kkwtlisknlrpedsgkytcr 222
 QY 63 ARNAGCLRADPPLSYVQR 81
 DB 223 vsnragalnaikyvdvlygr 241
 Db
 RESULT 8
 AAB65179
 ID AAB65179 standard; Protein; 504 AA.
 XX
 AC AAB65179;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Human PRO943 (UNC480) protein sequence SEQ ID NO:119.
 XX
 KW Human; secreted and transmembrane protein; PRO; cytosstatic;
 KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
 KW diagnostic assay.
 XX
 OS Homo sapiens.
 XX
 XX MO200073454-AL.
 PN
 XX 07-DEC-2000.
 PD
 XX
 PF 30-MAR-2000; 2000MO-US08439.
 XX
 XX 02-JUN-1999; 99MO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 07-JUL-1999; 99US-0143048.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 17-AUG-1999; 99US-0149396.
 PR 15-SEP-1999; 99MO-US21090.
 PR 15-SEP-1999; 99MO-US21547.
 PR 08-OCT-1999; 99MO-US21547.
 PR 30-NOV-1999; 99US-0158663.
 PR 01-DEC-1999; 99MO-US28301.
 PR 16-DEC-1999; 99MO-US30095.

PR 20-DEC-1999; 99MO-US30911.
 PR 05-JAN-2000; 2000MO-US00219.
 PR 06-JAN-2000; 2000MO-US00376.
 PR 11-FEB-2000; 2000MO-US03565.
 PR 18-FEB-2000; 2000MO-US04341.
 PR 22-FEB-2000; 2000MO-US04414.
 PR 24-FEB-2000; 2000MO-US04914.
 PR 02-FEB-2000; 2000MO-US05004.
 PR 15-MAR-2000; 2000MO-US05841.
 PR 20-MAR-2000; 2000MO-US07377.
 XX
 PA (GERTH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
 PI Ferrara N, Fong S, Gerber H, Gertlisen ME, Goddard A, Godowski PJ,
 PI Grimaldi CJ, Gurney AL, Kijavlin IJ, Napier MA, Pan J, Paoni NF,
 PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI,
 PI Zhang Z;
 XX
 DR WPI: 2001-032160/04.
 DR N-PSDB; AAF44130.
 XX
 PT PRO polynucleotides used to produce polypeptides used to target
 PT bioactive molecules such as toxins, radiolabels or antibodies, to
 PT specific cells, to cause targeted cell death -
 XX
 PS Claim 12; Fig 70; 935pp; English.
 XX
 CC The present invention describes human secreted and transmembrane PRO
 CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
 CC can be used for targeted delivery of bioactive molecules, such as
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
 CC sequences, and their fragments, can be used as hybridisation probes, in
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA
 CC and DNA. They may also be used to produce transgenic animals which are
 CC used to develop and screen therapeutically useful reagents. The PRO
 CC nucleotide and protein sequence can be used for tissue typing and in
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
 CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
 CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 504 AA:
 Query Match 28.4%; Score 125.5; DB 22; Length 504;
 Best Local Similarity 36.7%; Pred. No. 1.1e-06;
 Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;
 QY 3 LHCDVSGRPPAVTWKQSHORENLIMRPPQKGNVVTISIGOLVLYNARPDAGLYTCT 62
 DB 170 lkcvasghprpdltwmkd----dgaltrpeaaep---kkwtlisknlrpedsgkytcr 222
 QY 63 ARNAGCLRADPPLSYVQR 81
 DB 223 vsnragalnaikyvdvlygr 241
 Db
 RESULT 9
 AAB66264
 ID AAB66264 standard; Protein; 504 AA.
 XX
 AC AAB66264;
 XX
 DT 05-APR-2001 (first entry)
 XX
 DE Human MANGO 003 SEQ ID NO: 5.
 XX
 KW Membrane associated protein; secreted protein; human; mouse; rat;
 KW INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 354;
 KW TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;

KW haematopoietic disorder; neural disorder; hepatic disorder;
 KW neoplastic disease.
 XX Homo sapiens.
 OS
 XX WO200100673-A1.
 PN
 XX 04-JAN-2001.
 PD
 XX 29-JUN-2000; 2000WO-US18198.
 PF
 XX 30-JUN-1999; 99US-0345464.
 PR
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 XX Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;
 PI
 XX WPI: 2001-050128/06.
 DR
 XX N-PSDB; AAF27781.
 XX
 XX Isolated secreted or transmembrane proteins are used for diagnosis and
 PT treatment of neoplastic and haematopoietic disorders e.g. T cell
 PT disorders, cancer and tumours -
 XX
 XX Claim 9; Page 216-217; 294pp; English.
 PS
 XX The present invention provides the protein and coding sequences for a
 CC number of membrane associated and secreted proteins from human, mouse and
 CC rat. The proteins are designated INTERCEPT 340, MANO 003, MANO 347,
 CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
 CC involved in signal transduction and the sequences can be used in the
 CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
 CC and haematopoietic disorders.
 CC
 XX Sequence 504 AA;
 SQ
 XX
 XX Query Match 28.4%; Score 125.5; DB 22; Length 504;
 XX Best Local Similarity 36.7%; Pred. No. 1.1e-06;
 XX Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;
 OY 3 LHCDVSGRPPTAVTWKQSHORENLIMRPDMGNVVTISIGLVLYNARPDAGLYTCT 62
 Db 170 lkcvasghprpdltmkd----dgaltlrpeaaepr---kkwtlslknlrpdsgkytcr 222
 OY 63 ARNAAGLLRADPPLSVYOR 81
 Db 223 vsnragainatykvdivigr 241
 RESULT 10
 AAM38643
 ID AAM38643 standard; Protein; 608 AA.
 XX
 AC AAM38643;
 XX
 XX 22-OCT-2001 (first entry)
 DT
 XX Human polypeptide SEQ ID NO 1788.
 DE
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 XX Homo sapiens.
 OS
 XX WO2001053312-A1.
 PN
 XX 26-JUL-2001.
 PD

PF 26-DEC-2000; 2000WO-US34263.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSEQ-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;
 XX
 XX WPI: 2001-442253/47.
 DR
 XX N-PSDB; AAI57799.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 XX
 XX Example 3; SEQ ID NO 1788; 10078pp; English.
 PS
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX Sequence 608 AA;
 SQ
 XX
 XX Query Match 28.4%; Score 125.5; DB 22; Length 608;
 XX Best Local Similarity 36.7%; Pred. No. 1.3e-06;
 XX Matches 29; Conservative 9; Mismatches 34; Indels 7; Gaps 2;
 OY 3 LHCDVSGRPPTAVTWKQSHORENLIMRPDMGNVVTISIGLVLYNARPDAGLYTCT 62
 Db 157 lkcvasghprpdltmkd----dgaltlrpeaaepr---kkwtlslknlrpdsgkytcr 209
 OY 63 ARNAAGLLRADPPLSVYOR 81
 Db 210 vsnragainatykvdivigr 228
 RESULT 11
 AAG62783
 ID AAG62783 standard; Protein; 545 AA.
 XX
 AC AAG62783;
 XX
 XX 17-SEP-2001 (first entry)
 DT
 XX Amino acid sequence of an interleukin-1 receptor-like polypeptide.
 DE
 XX Expressed sequence tag; EST; thyroid gland; sepsis; endotoxin shock;
 KW interleukin-1 receptor-like polypeptide; auto-immune disease; cancer;
 KW IL-1 receptor-like polypeptide; arthritis; inflammatory disease.
 KW
 XX Homo sapiens.
 OS

```

XX Key Location/Qualifiers
FH Domain 120..152
FT /note="IL-1 receptor precursor domain"
FT Domain 279..302
FT /note="tyrosine kinase class III receptor domain"
FT Domain 374..397
FT /note="tyrosine kinase class III receptor domain"
FT Domain 470..493
FT /note="tyrosine kinase class III receptor domain"
FT Domain 512..544
FT /note="IL-1 receptor precursor domain"
XX
XX MO200153515-A1.
XX
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000MO-US35263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 05-MAY-2000; 2000US-0565522.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Boyle BJ, Mize NK, Arterburn MC, Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-451912/48.
XX DR N-PSDB; AAH42133.
XX
XX Isolated polypeptide with interleukin-1 receptor-like activity for
XX treatment of disorders involving sepsis, endotoxic shock, arthritis,
XX auto-immune disease, inflammatory disease or cancer -
XX
XX Claim 10; Page 133-135; 143pp; English.
XX
XX The present sequence represents human interleukin-1 (IL-1) receptor-like
XX polypeptide. The cDNA sequence is isolated from a cDNA library derived
XX from human thyroid gland. IL-1 receptor-like polypeptides are useful for
XX treating a subject in need of interleukin-1 receptor-like activity. They
XX may be used for the prevention and treatment of disorders involving
XX sepsis, endotoxic shock, arthritis, auto-immune disease, inflammatory
XX disease or cancer.
XX
XX Sequence 545 AA:
SQ

```

Query Match 27.4%; Score 121; DB 22; Length 545;
 Best Local Similarity 32.1%; Pred. No. 4, 1e-06;
 Matches 25; Conservative 17; Mismatches 32; Indels 4; Gaps 2;

```

QY 3 LHCDSGRPPAVTWKQSHORENLIIRPDQMGVNVVTSIGOLVLYNARPEADAGLYTCT 62
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 470 lhcvalgyvpkpetlwemphsllstask-erthgseqhlqgtlvtqnpqtadsqlykct 528
QY 63 ARNAGCLRADPPLSVQ 80
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 529 aknp1g---sdyaaty1q 543

```

RESULT 12
 AAG62782
 ID AAG62782 standard; Protein; 579 AA.
 XX
 AC AAG62782;
 XX
 DT 17-SEP-2001 (first entry)
 XX
 DE Amino acid sequence of an interleukin-1 receptor-like polypeptide.
 XX
 KW Expressed sequence tag; EST; thyroid gland; sepsis; endotoxic shock;
 KW interleukin-1 receptor-like polypeptide; auto-immune disease; cancer;
 KW IL-1 receptor-like polypeptide; arthritis; inflammatory disease.

```

XX OS Homo sapiens.
XX PN WO200153515-A1.
XX
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000MO-US35263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 05-MAY-2000; 2000US-0565522.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Boyle BJ, Mize NK, Arterburn MC, Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-451912/48.
XX DR N-PSDB; AAH42132.
XX
XX Isolated polypeptide with interleukin-1 receptor-like activity for
XX treatment of disorders involving sepsis, endotoxic shock, arthritis,
XX auto-immune disease, inflammatory disease or cancer -
XX
XX Example 2; Page 113; 143pp; English.
XX
XX The present sequence represents human interleukin-1 (IL-1) receptor-like
XX polypeptide. The cDNA sequence is isolated from a cDNA library derived
XX from human thyroid gland. IL-1 receptor-like polypeptides are useful for
XX treating a subject in need of interleukin-1 receptor-like activity. They
XX may be used for the prevention and treatment of disorders involving
XX sepsis, endotoxic shock, arthritis, auto-immune disease, inflammatory
XX disease or cancer.
XX
XX Sequence 579 AA:
SQ

```

Query Match 27.4%; Score 121; DB 22; Length 579;
 Best Local Similarity 32.1%; Pred. No. 4, 4e-06;
 Matches 25; Conservative 17; Mismatches 32; Indels 4; Gaps 2;

```

QY 3 LHCDSGRPPAVTWKQSHORENLIIRPDQMGVNVVTSIGOLVLYNARPEADAGLYTCT 62
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 504 lhcvalgyvpkpetlwemphsllstask-erthgseqhlqgtlvtqnpqtadsqlykct 562
QY 63 ARNAGCLRADPPLSVQ 80
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 563 aknp1g---sdyaaty1q 577

```

RESULT 13
 AAW74152
 ID AAW74152 standard; Protein; 1257 AA.
 XX
 AC AAW74152;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Human IL cell adhesion molecule.
 XX
 KW Human IL cell adhesion molecule; L1CAM; neurite growth;
 KW nervous system development; nerve regeneration;
 KW neuronal cell cohesive interaction.
 XX
 OS Homo sapiens.
 XX
 PN US5872225-A.
 XX
 PD 16-FEB-1999.
 XX
 PF 18-NOV-1994; 94US-0341843.
 XX
 PR 26-JUN-1992; 92US-0904991.

PR 18-NOV-1994; 94US-0341843.
 XX
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 XX
 PI Lemmon V;
 XX
 DR WPI; 1999-166719/14.
 DR N-PSDB; AAX01598.
 XX
 PT Human LI cell adhesion molecule - supports neurite outgrowth and is
 XX involved in nervous system development and repair
 PS Claim 1; Fig 3; 45pp; English.
 XX
 CC This sequence is the human LI cell adhesion molecule (LICAM) of the
 CC invention. LICAM supports growth of neurites in vitro and is involved in
 CC development of the human nervous system and in nerve regeneration. It is
 CC useful in vivo and in vitro experiments on nerve growth and
 CC regeneration. LICAM mediates cohesive interactions of neuronal cells to
 CC each other and to extracellular matrix.
 XX
 SQ Sequence 1257 AA;

Query Match 27.4%; Score 121; DB 20; Length 1257;
 Best Local Similarity 41.2%; Pred. No. 1.1e-05;
 Matches 33; Conservative 5; Mismatches 36; Indels 6; Gaps 2;

OY 1 ASLHCDSGRRPPAVTWEKOSHORENLIMRPDQMGVWVTSIGQLVLYNARPEADAGLYTCR 60
 Db 350 arldcgvqgvpqevtrringipveel--akdqkyr---lqgalllnvqpsdlnvtg 403
 OY 61 CTARNAGLLRADFPISVQ 80
 Db 404 cearnthgillanaylyvvg 423

RESULT 14
 AAY13567
 ID AAY13567 standard; Protein; 434 AA.
 XX
 AC AAY13567;
 XX
 DT 30-JUL-1999 (first entry)
 XX
 DE Human Robo 2 polypeptide.
 XX
 KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
 KW modulation; nerve cell function.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 285
 FT /label= unknown
 FT /note= "encoded by GTN"
 FT Misc-difference 396
 FT /label= unknown
 FT /note= "encoded by NTR"
 XX
 XX WO9925833-A1.
 XX
 XX PD 27-MAY-1999.
 XX
 XX PF 13-NOV-1998; 98WO-US24327.
 XX
 XX PR 14-NOV-1997; 97US-0065543.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Goodman C, Kidd T, Mitchell KJ, Russell C, Tear G;
 XX WPI; 1999-338008/28.
 DR

DR N-PSDB; AAX55771.
 XX
 PT Modulation of Robo-Comm polypeptide interactions
 XX
 PS Disclosure; Page 49-50; 56pp; English.
 XX
 CC The invention relates to a method for modulating the amount of Comm
 CC (commissureless) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo:Comm
 CC interactions. This is particularly useful for modulating nerve cell
 function.
 XX
 SQ Sequence 434 AA;

Query Match 27.3%; Score 120.5; DB 20; Length 434;
 Best Local Similarity 32.3%; Pred. No. 3.5e-06;
 Matches 30; Conservative 12; Mismatches 36; Indels 15; Gaps 3;

OY 5 CDVSGRRPPAVTWEKOSHORENLIM--RPDQMGVWVTSIGQLVLYNARPEADAGLYTCR 62
 Db 13 cckgppqpvfvgkqsg--nllfpnqpgqpnstcspsptgdlitlnqrsdagyicq 70
 OY 63 ARNAGILRA-----DFPLSVQREPA 84
 Db 71 alvagsilakaglevrldvldtrppillqgpa 103

RESULT 15
 AAY08405
 ID AAY08405 standard; Protein; 434 AA.
 XX
 AC AAY08405;
 XX
 DT 24-JUL-1999 (first entry)
 XX
 DE Human partial ROBO2 protein.
 XX
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
 KW cell morphology; screening assay.
 XX
 OS Homo sapiens.
 XX
 FH
 FT WO9920764-A1.
 FT 29-APR-1999.
 XX
 XX PF 20-OCT-1998; 98WO-US22164.
 XX
 XX PR 14-NOV-1997; 97US-0971172.
 XX
 XX PR 20-OCT-1997; 97US-0062921.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Goodman CS, Kidd T, Mitchell KJ, Tear G;
 XX WPI; 1999-312615/26.
 XX
 XX DR N-PSDB; AAX57254.
 XX
 XX Robo polypeptides, a new immunoglobulin superfamily member
 XX Claim 1; Page 72-73; 80pp; English.
 XX
 CC This invention describes novel Robo (roundabout) polypeptides, involved
 CC in nerve guidance which have been isolated from *Drosophila* sp.,
 CC C. elegans, human and murine samples. The products of the invention can
 CC be used to raise anti-Robo antibodies, which can be used to modulate cell
 CC function or morphology. The Robo polynucleotides and fragments are useful

CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.

Sequence 434 AA;

Query Match	Score	DB	Length
27.38;	120.5;	20;	434;
Best [local] similarity	37.38;	Pred NO	3 5a-06;

Best Local Similarity 32.3%; Pred. No. 3.5e-06;
Matches 30; Conservative 12; Mismatches 36; Indels 15; Gaps 3;

```

OY      5 CDVSGRRPRAVWEKSHORENLM--RPDOMGYNVVTISICGLVLYNARPEDACLYTCT 62
      1 : 1 1 1 1 : 1 1 : 1 : 1 : 1 1 1
Db     13 cctkgnpqpvafwqkegsq-nllfpnqpqqnsrscsvsptdltltitqrsdagyyicq 70

```

```
l : | | ||| |: | ||: :: | | | : | | : ||| | |  
celkgnpq pavfwqkegsq--nllfpnqpqpnsrcsvsp tgdltltlnigrsdagyy lcg 70
```

QY 63 ARNAGGLRA-----DFPLSVQREPA 84

Db 71 altvagsllakaglevtvlttdrpplllgpa 103

Search completed: February 26, 2002, 01:26:43
Job time: 317 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:20 ; Search time 144.8 seconds
(without alignments)
53.539 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_351
Perfect score: 314
Sequence: 1 CLPDVQACTGPTSPHLVLMH.....RGCDGARGFETYEACQAC 53

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.5	36.5	1743	5	Q9XW5 caenorhabd
2	110.5	35.2	3060	5	Q9YAV4
3	105	33.4	3198	5	Q9868
4	104	33.1	195	11	Q9D808
5	94.5	30.1	151	4	P78491
6	94.5	30.1	335	5	O44622
7	94	29.9	113	5	O9V508
8	94	29.9	607	11	O99K32
9	92.5	29.5	922	5	O21418
10	92	29.3	74	11	O60495
11	92	29.3	160	11	O90278
12	92	29.3	484	4	O13793
13	92	29.3	523	4	O14594
14	92	29.3	547	4	O13764
15	92	29.3	751	11	O60709
16	92	29.3	763	11	O61482
17	92	29.3	770	6	O9TUI0
18	91.5	29.1	1572	5	O44938
19	91	29.0	249	5	O16701

20	91	29.0	1965	5	O61893	O61893 caenorhabd
21	90.5	28.8	2174	5	O9C0R0	O9C0R0 drosophila
22	90	28.7	342	13	P70004	P70004 xenopus lae
23	90	28.7	737	13	O93279	O93279 fugu rubrip
24	89.5	28.5	67	11	O9QW86	O9QW86 mesocricetu
25	89.5	28.5	144	11	O9QW87	O9QW87 mesocricetu
26	89.5	28.5	2167	5	O76840	O76840 caenorhabd
27	88	28.0	55	5	P81902	P81902 bombyx mori
28	88	28.0	751	13	O9D37	O9D37 gallus gall
29	86.5	27.5	349	11	O9DBY9	O9DBY9 mus musculu
30	86	27.4	747	13	O91963	O91963 xenopus lae
31	84	26.8	133	6	O9BD11	O9BD11 macaca mula
32	83.5	26.6	97	5	O9VOT1	O9VOT1 drosophila
33	83	26.4	58	5	O9TXD3	O9TXD3 fasciola he
34	83	26.4	113	5	O9V5S6	O9V5S6 drosophila
35	83	26.4	616	11	O99K31	O99K31 mus musculu
36	83	26.4	2657	11	O88493	O88493 mus musculu
37	82.5	26.3	352	11	O70160	O70160 cavia porce
38	82	26.1	1391	5	P78492	P78492 homo sapien
39	82	26.1	1391	5	O19021	O19021 caenorhabd
40	82	26.1	1522	5	O22685	O22685 caenorhabd
41	80	25.5	507	11	O9D3K4	O9D3K4 mus musculu
42	80	25.5	507	11	O99J04	O99J04 mus musculu
43	80	25.5	838	5	O27422	O27422 caenorhabd
44	80	25.5	838	5	O18761	O18761 caenorhabd
45	80	25.5	1208	5	O23456	O23456 caenorhabd

ALIGNMENTS

RESULT 1
ID Q9XW5 PRELIMINARY; PRT: 1743 AA.
AC Q9XW5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Y43F8B.3 PROTEIN.
GN Y43F8B.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Ainscough R.;
RL Submitted (OCF-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Garner A., Green P., Hawkins J., Kirschen J., Lalster N., Johnston L.,
RA Jones M., Kershaw J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Lightning J., Percy C., Ritten L., Roopra A., Saunders D., Showkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spoat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AL032623; CAA21511.1; -.
DR HSSP: P05067; ITAW.
DR InterPro: IPR002899; EB.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 16.
DR PRINTS: PR00759; BASICPRASE.
DR SMART: SM00131; KU; 15.
DR SMART; SM00289; WRI; 10.

DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 15.
 KW Serine protease inhibitor.
 SO SEQUENCE 1743 AA; 19111 MM; B93C163556433C2A CRC64;

Query Match 36.5%; Score 114.5; DB 5; Length 1743;
 Best Local Similarity 39.3%; Pred. No. 3.4e-06;
 Matches 22; Conservative 8; Mismatches 21; Indels 5; Gaps 2;

OY 3 PDVQACTP---TSPHLY-LMHYPQGGCMTFARCCDGAARCFEYACQAC 53
 Db 71 PSQSGCTLPKQIGTSPRIPIPRMYNVPKRCLEFYWGSCGNGNMFQTCQSTC 126

RESULT 2
 O9AV4 PRELIMINARY; PRT; 3060 AA.
 AC O9AV4; O9AV3; (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG1540 PROTEIN.
 GN PPN OR CG1540 OR CG18436.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Anell J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Beus P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Evans G.S., Fink D., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fester A., Gabori A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagham C.,
 RA Jalili M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklow G., Misha N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.C.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 CC EMBL; AE003765; AAF56794.2; -.
 CC EMBL; AE003765; AAF56795.2; -.
 DR HSSP; P12111; 2KNT.
 DR FLYBASE; FBgn0003137; PPN.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR002221; MAP.
 DR Pfam; PF00047; Ig_3.
 DR Pfam; PF00014; Kunitz_BPTI; 12.
 DR Pfam; PF00090; tsp_1; 5.
 DR Pfam; PF00095; wap; 1.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00408; IGG2; 3.
 DR SMART; SM00131; KU; 12.
 DR SMART; SM00209; TSP1; 7.
 DR SMART; SM00217; WAP; 1.
 DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 11.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 12.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS50092; TSP1_3.
 DR Alternative splicing; Serine protease inhibitor.
 KW Alternative splicing; Serine protease inhibitor.
 FT VARSPLIC 2803 2803
 FT VARSPLIC 2844 2854
 FT VARSPLIC 2855 3060
 FT VARSPLIC 3060 AA; 331579 MM; ACA31D3BE558C7C0 CRC64;
 SO SEQUENCE

Query Match 35.2%; Score 110.5; DB 5; Length 3060;
 Best Local Similarity 40.0%; Pred. No. 1.9e-05;
 Matches 22; Conservative 6; Mismatches 24; Indels 3; Gaps 2;

OY 1 CLPDVQA--CTGPTSHLYLMHDPQGGCMTFARCCDGAARCFEYACQAC 53
 Db 2278 CLLPVATGRGNP-SYHERRWYDDEAGNCVSFTYAGCGNNGNFRSFACYNQC 2331

RESULT 3
 Q9U8G8 PRELIMINARY; PRT; 3198 AA.
 AC Q9U8G8; (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE LACUNIN PRECURSOR.
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
 NCBI_TaxID=7130;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99457716; PubMed=10528409;
 RA Nairi J.B., Marlos R., Walden K.K., Lampe D.J., Robertson H.M.;
 RT "Expression of lacunin, a large multidomain extracellular matrix
 RT protein, accompanies morphogenesis of epithelial monolayers in *Manduca*
 RT sexta";
 RL Insect Biochem. Mol. Biol. 29:883-897(1999).
 CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 CC EMBL; AF078161; AAF04457.1; -.
 DR HSSP; P12111; 2KNT.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR002223; Kunitz_BPTI.

DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00047; Ig 2.
 DR Pfam: PF00014; Kunitz_BPTI; 9.
 DR Pfam: PF00095; Wap; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00408; IGC2; 2.
 DR SMART: SM00131; KU; 10.
 DR SMART: SM00209; TSP1; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 8.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 10.
 DR PROSITE: PS50092; TSP1; 1.
 KW Serine protease inhibitor; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 SQ SEQUENCE 3198 AA; 349366 MW; AB4ACD459C0D9134 CRC64;

Query Match 33.4%; Score 105; DB 5; Length 3198;
 Best Local Similarity 40.0%; Pred. No. 0.0001;
 Matches 18; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 9 TGPTSPHLVLMHYDPQRCGCMTPFARGCDGAARGFETYEACQAC 53
 DB 2269 SGPCTESLMRWFPYDPSSDSCSOFTYCGCGDNRRFETRDCESRC 2313

RESULT 4
 Q9D808 PRELIMINARY; PRT; 195 AA.

AC Q9D808: 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE SERINE PROTEASE INHIBITOR, KUNITZ TYPE 2.
 GN SPINT2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guerninich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombearts P.,
 RA Noriote P., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Teyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuk S.,
 RA Hayshtizaki Y.,
 FT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AK007793; BAB25258.1; -.
 DR MGI: MGI:138031; Spint2.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.

DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
 KW Serine protease inhibitor.
 SQ SEQUENCE 195 AA; 21705 MW; 1ABD78CEFF1175DE1 CRC64;

Query Match 33.1%; Score 104; DB 11; Length 195;
 Best Local Similarity 35.8%; Pred. No. 8.9e-06;
 Matches 19; Conservative 9; Mismatches 23; Indels 2; Gaps 1;

QY 1 CLPDVACTGPTSPHLVLMHYDPQRCGCMTPFARGCDGAARGFETYEACQAC 53
 DB 76 CVP-KAVTGPCRAAFPRWYDYTEKNSCISFTYGGCGKNKNSYLSEACMOMC 126

RESULT 5
 P78491 PRELIMINARY; PRT; 151 AA.

AC P78491: 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE ALPHA-1-MICROGLOBULIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91214554; PubMed=1708673;
 RA Vetr H., Gebhard W.;
 RT "Structure of the human alpha 1-microglobulin-bikunin gene";
 RL Bio. Chem. Hoppe-Seyler 371:1185-1196(1990).
 CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: X54818; CA38587.1; -.
 DR EMBL: X54816; CA38587.1; JOINED.
 DR HSSP: P02760; IBIK.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 KW Serine protease inhibitor.
 FT NON_TER 1 1
 SQ SEQUENCE 151 AA; 16542 MW; 88F400C5ECA19831 CRC64;

Query Match 30.1%; Score 94.5; DB 4; Length 151;
 Best Local Similarity 33.3%; Pred. No. 0.00012;
 Matches 18; Conservative 8; Mismatches 23; Indels 5; Gaps 1;

QY 5 VOACT-----GPTSPHLVLMHYDPQRCGCMTPFARGCDGAARGFETYEACQAC 53
 DB 83 VAAACNPIYRGPRAFIOLMAFDAYKGCVLFPYGGCGNGNKFYSEKCREYC 136

RESULT 6
 O44622 PRELIMINARY; PRT; 335 AA.

AC O44622: 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE KIID12.7 PROTEIN.
 GN KIID12.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;

RA Sutton G., Wolman J.R., Iandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abbayanti A., An H.-J., Andrews-Pfankoch C., Baldwin D.,

RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases
DR EMBL; BC005490; AAH05490.1; -.
ET NON TER 1 1

SO SEQUENCE 607 AA: 68391 MW: BF802214CBA7D172 CRC64:

Query Match 29.9%; Score 94; DB 11; Length 607;
Best Local Similarity 37.5%; Pred. No. 0.00052;
Matches 18; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

Oy 6 QACTGPTSPHLVLMHYDPQRCGCMTPFARGCDGARGFETYACQOAC 53
Db 131 QACTGPRAMISRWYEDYEGKCAPEFYGGCGGNRNPFDEECYMAVC 178

RESULT 9

O21418 PRELIMINARY: PRT: 922 AA.
AC Q21418;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE K10D3.4 PROTEIN.
GN K10D3.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RA McMurray A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RP MEDLINE=94150718; PubMed=7906398;
RX Wilson R., Alnscough R., Anderson K., Baynes C., Berks M., Coulson A., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L., Craiton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latelle P., Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R., Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";
RL Nature 368:32-38(1994).
CC -1 SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: Z75545; CAA9886.1; -;
DR HSSP: P10646; ITRX.
DR InterPro: IPR002899; EB.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF01683; EB; 1.
DR Pfam: PF00014; Kunitz_BPTI; 5.
DR PRINTS: PR00759; BASICPASE.
DR SMART: SM00131; KU; 5.
DR SMART: SM00289; BPTI_KUNITZ_1; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_2; 5.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 5.
KW Serine protease inhibitor.
SQ SEQUENCE 922 AA: 99995 MW: 0CCBCC0AE2524CFD CRC64:

Query Match 29.5%; Score 92.5; DB 5; Length 922;
Best Local Similarity 38.5%; Pred. No. 0.0012;
Matches 20; Conservative 4; Mismatches 25; Indels 3; Gaps 1;

Oy 2 LPDVACTGPTSPHLVLMHYDPQRCGCMTPFARGCDGARGFETYACQOAC 53
Db 413 LPREGNGCTYSNR---WTFNAKTGNCCEFIYSGCGNANNFETYECCDYC 461

RESULT 10
O60495 PRELIMINARY: PRT: 74 AA.

AC 060495;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN (FRAGMENT).
GN APP.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Hystriognathii; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RN SEQUENCE FROM N.A.
RA Beck M., Mueller D., Bigl V.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1 FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN G(O) (BY SIMILARITY).
CC -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1 SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: X99198; CAA67589.1; -;
DR HSSP: P05067; ITRM.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPASE.
DR SMART: SM00131; KU; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor.
FT NON_TER 1 74 74
SQ SEQUENCE 74 AA: 8057 MW: 29AB3407730814D9 CRC64:

Query Match 29.3%; Score 92; DB 11; Length 74;
Best Local Similarity 37.5%; Pred. No. 0.00012;
Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Oy 6 QACTGPTSPHLVLMHYDPQRCGCMTPFARGCDGARGFETYACQOAC 53
Db 5 QACTGPRAMISRWYEDYEGKCAPEFYGGCGGNRNPFDEECYMAVC 52

RESULT 11
O60278 PRELIMINARY: PRT: 160 AA.
AC Q60278;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE AMYLOID PRECURSOR PROTEIN (FRAGMENT).
OS Cavia sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Hystriognathii; Cavidae; Cavia.
OX NCBI_TaxID=10143;
RN [1]
RN SEQUENCE FROM N.A.
RA Tissue-Lens;
RC Tissue-Lens;
RA Frederikse P.H., Carper D., Farnsworth J.P., Zigler J.S.;
RT "Pilon and Alzheimer precursor protein expression in a hereditary guinea pig cataract.";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF197164; AAF08934.1; -;
DR HSSP: P05067; IAP.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPASE.
DR SMART: SM00131; KU; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor.
FT NON_TER 1 1 1


```

DR InterPro: IPR001868; A4_APP.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF02177; A4_EXTRA; 1.
DR Pfam: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00006; A4_EXTRA; 1.
DR SMART: SM00006; A4_EXTRA; 1.
DR SMART: SM00131; KU; 1.
DR PROSITE: PS00319; A4_EXTRA; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
DR Serine protease inhibitor.
SQ SEQUENCE 547 AA; 62135 MW; D81B1E0FF2D2D070 CRC64;

Query Match          29.3%; Score 92; DB 4; Length 547;
Best Local Similarity 37.5%; Pred. NO. 0.00085;
Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Oy 6 QACTGPTSPHLVLMHYDPQRGCMTPFARGCDGAARGFETYEACQOAC 53
Db 294 QAETGPCRAMISRMYFDVTEGKCAPFFYGGCGGNRNRFTEYCAVAC 341

RESULT 15
O60709 PRELIMINARY: PRT: 751 AA.
AC O60709.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE AMYLOID BETA (A4) PRECURSOR-LIKE PROTEIN 2 (AMYLOID-LIKE PROTEIN 2,
DE ISOFORM 751).
CN APLP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ICR:
RA Slunt H.H., Von Koch C., Sisodia S.S.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-229 FROM N.A.
RC STRAIN-ICR:
RX MEDLINE=94132029; PubMed=8300594;
RA Slunt H.H., Thinakaran G., Von Koch C., Lo A.C., Tanzi R.E.,
RA Sisodia S.S.;
RT "Expression of a ubiquitous, cross-reactive homologue of the mouse
RT beta-amyloid precursor protein (APP).";
RL J. Biol. Chem. 269:2637-2644(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ICR:
RX MEDLINE=94350957; PubMed=8071334;
RA Thinakaran G., Sisodia S.S.;
RT "Amyloid precursor-like protein 2 (APLP2) is modified by the addition
RT of chondroitin sulfate glycosaminoglycan at a single site.";
RL J. Biol. Chem. 269:22099-22104(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-ICR:
RA Thinakaran G.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: U15571; AA50603.1; -.
DR HSSP: P05067; IAMP.
DR MGD: MGI:88047; APLP2.
DR InterPro: IPR001868; A4_APP.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF02177; A4_EXTRA; 1.
DR Pfam: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00203; AMYLOIDA4.

```

```

DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00006; A4_EXTRA; 1.
DR SMART: SM00131; KU; 1.
DR PROSITE: PS00319; A4_EXTRA; 1.
DR PROSITE: PS00320; A4_INTRA; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 1.
DR Serine protease inhibitor.
SQ SEQUENCE 751 AA; 85247 MW; D183F6A4AD17BC04 CRC64;

Query Match          29.3%; Score 92; DB 11; Length 751;
Best Local Similarity 33.3%; Pred. NO. 0.0012;
Matches 16; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

Oy 6 QACTGPTSPHLVLMHYDPQRGCMTPFARGCDGAARGFETYEACQOAC 53
Db 313 EAMTGCRVAMPMPWFDLSKRCVRFYGGCGGNRNRFESDYCAVAC 360

```

Search completed: February 26, 2002, 01:49:22
Job time: 1271 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:40 ; Search time 45.78 Seconds

(without alignments)
42.447 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_351
Perfect score: 314
Sequence: 1 CLPDVQACTGPTSPHLVLMH.....RGCDGARGEFYEACQAC 53

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProtC_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	33.1	252	1	SPT2_MOUSE
2	94.5	30.1	352	1	AMBP_MOUSE
3	94	29.9	235	1	TPP2_HUMAN
4	94	29.9	770	1	A4_MOUSE
5	92	29.3	76	1	A4_MACMU
6	92	29.3	87	1	A4_MACEA
7	92	29.3	751	1	A4_SAISSC
8	92	29.3	763	1	APP2_HUMAN
9	92	29.3	763	1	APP2_MOUSE
10	92	29.3	770	1	A4_HUMAN
11	92	29.3	770	1	A4_RAT
12	91.5	29.1	123	1	IATR_SHEEP
13	91.5	29.1	352	1	AMBP_MOUSE
14	90.5	28.8	346	1	AMBP_MERUN
15	89.5	28.5	349	1	AMBP_MESAU
16	88	28.0	58	1	ISIK_HELPO
17	88	28.0	252	1	SPT2_HUMAN
18	87.5	27.9	61	1	IVBI_VIPPA
19	86.5	27.5	123	1	IATR_HORSE
20	86.5	27.5	349	1	AMBP_MOUSE
21	85.5	27.2	337	1	AMBP_PIG
22	85	27.1	1416	1	YN81_CAREL
23	85	27.1	3137	1	CA36_CHICK
24	83.5	26.6	62	1	IVBI_ERIMA
25	82.5	26.3	349	1	AMBP_RAT
26	82	26.1	100	1	BPT1_BOVIN
27	82	26.1	100	1	BPT2_BOVIN
28	82	26.1	122	1	UPT1_PIG
29	82	26.1	265	1	TKD1_SHEEP
30	81	25.8	60	1	IBPS_BOVIN
31	80	25.5	60	1	IVB2_DABRU
32	80	25.5	230	1	TFP2_MOUSE
33	80	25.5	507	1	SPT1_MOUSE

34	79.5	25.3	64	1	SPT3_HUMAN	P49223 homo sapien
35	79	25.2	513	1	SPT1_HUMAN	O43278 homo sapien
36	78.5	25.0	65	1	IVB3_VIPPA	P00992 vipera ammo
37	78	24.8	164	1	TKD1_BOVIN	Q28201 bos taurus
38	77	24.5	3176	1	CA36_HUMAN	P12111 homo sapien
39	76.5	24.4	62	1	IP52_ANESU	P10280 anemolia su
40	76	24.2	57	1	IVB2_HEMMA	P00985 hemachatus
41	76	24.2	57	1	IVB2_DENAN	P00982 dendroaspis
42	76	24.2	133	1	EPPI_HUMAN	O59925 homo sapien
43	75.5	24.0	67	1	IBPC_BOVIN	P00976 bos taurus
44	75	23.9	57	1	IVBK_DENNO	P00981 dendroaspis
45	75	23.9	59	1	IVBE_DENNO	P00984 dendroaspis

ALIGNMENTS

RESULT	1	STANDARD	PRT	252 AA.
AC	SPT2_MOUSE			
DT	Q9WU03; Q9WU04; Q9WU05;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DE	KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR ACTIVATOR INHIBITOR TYPE 2) (HA1-2).			
GN	SPTNT2 OR HA12.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			
RP	STRAIN=BAIB/C;			
RX	MEDLINE=99160423; PubMed=10049781;			
RA	Itoh H., Kataoka H., Hamasuna R., Kitamura N., Koono M.;			
RT	"Hepatocyte growth factor activator inhibitor type 2 lacking the first			
RT	kunitz-type serine proteinase inhibitor domain is a predominant			
RT	product in mouse but not in human."			
RL	Biochem. Biophys. Res. Commun. 255:740-748(1999).			
CC	-1- FUNCTION: INHIBITOR OF HGF ACTIVATOR.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE			
CC	PRODUCED BY ALTERNATIVE SPLICING.			
CC	-1- TISSUE SPECIFICITY: ISOFORM 2 IS MORE PREDOMINANTLY EXPRESSED THAN			
CC	ISOFORM 1.			
CC	-1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 2 BPT1/KUNITZ INHIBITOR DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF099016; AAD22172.1; -			
DR	EMBL; AF099019; AAD22173.1; -			
DR	EMBL; AF099020; AAD22174.1; -			
DR	HSSP; P05067; ITAM.			
DR	MCD; MGI:1338031; SpInt2.			
DR	InterPro; IPR002223; Kunitz_BPT1.			
DR	Pfam; PF00014; Kunitz_BPT1; 2.			
DR	PRINTS; PR00759; BASICPTASE.			
DR	SMART; SM00131; KU; 2.			
DR	PROSITE; PS00280; BPT1_KUNITZ_1; 2.			
DR	PROSITE; PS50279; BPT1_KUNITZ_2; 2.			
KW	Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;			
KW	Signal; Alternative splicing.			
FT	CHAIN 1 27			
FT	FT 28 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.			
FT	DOMAIN 28 197 EXTRACELLULAR (POTENTIAL).			

```

FT TRANSMEM 198 218 POTENTIAL.
FT DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
FT DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
FT DISULFID 38 88 BY SIMILARITY.
FT DISULFID 47 71 BY SIMILARITY.
FT DISULFID 63 84 BY SIMILARITY.
FT AC1_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 133 183 BY SIMILARITY.
FT DISULFID 142 166 BY SIMILARITY.
FT DISULFID 158 179 BY SIMILARITY.
FT AC1_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 37 93 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPPLIC 114 128 PRKOSAEEDLSAEIEN -> CVELSVALLFLFYA (IN ISOFORM 3).
FT VARSPPLIC 129 252 MISSING (IN ISOFORM 3).
FT SEQUENCE 252 AA; 27914 MW; B2FF4B86924D4F8F CRC64;

```

```

Query Match 33.1%; Score 104; DB 1; Length 252;
Best Local Similarity 35.8%; Pred. No. 4.6e-06;
Matches 19; Conservative 9; Mismatches 23; Indels 2; Gaps 1;

```

```

OY 1 CLPDVQACTGPTSPHLVIMHYDPORGCMTEPPARCCDGAARGFETYEACQOAC 53
DB 133 CVP--KAVTGPCRAAFPRMYDYDEKNSCISFTYGGCRGNKNSYLSQACMCQHC 183

```

```

RESULT 2
AMBP_HUMAN STANDARD; PRT; 352 AA.
ID P02760; P02759; P00977;
AC 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN (PROTEIN HC)
DE (COMPLEX-FORMING GLYCOPROTEIN HETEROGENEOUS IN CHARGE); INNER-ALPHA-
DE TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
GN AMP OR ITI OR HCP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91214554; PubMed=1708673;
RA Vetr H., Gebhard W.;
RT "Structure of the human alpha 1-microglobulin-bikunin gene.";
RL Biol. Chem. Hoppe-Seyler 371:1185-1196(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87040757; PubMed=2430261;
RA Kaumeyer J.F., Polazzi J.O., Kotick M.P.;
RT "The mRNA for a proteinase inhibitor related to the HI-30 domain of
RT inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin
RT (protein HC).";
RL Nucleic Acids Res. 14:7839-7850(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90336621; PubMed=1696200;
RA Diarra-Mehrpour M., Bourguignon J., Sesboue R., Saller J.P.,
RA Leveillard T., Martin J.P.;
RT "Structural analysis of the human inter-alpha-trypsin inhibitor
RT light-chain gene.";
RL Eur. J. Biochem. 191:131-139(1990).
RN [4]
RP SEQUENCE OF 1-220 FROM N.A.
RX MEDLINE=86312901; PubMed=2428011;
RA Traboni C., Cortese R.;
RN

```

```

FT TRANSMEM 198 218 POTENTIAL.
FT DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
FT DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
FT DISULFID 38 88 BY SIMILARITY.
FT DISULFID 47 71 BY SIMILARITY.
FT DISULFID 63 84 BY SIMILARITY.
FT AC1_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 133 183 BY SIMILARITY.
FT DISULFID 142 166 BY SIMILARITY.
FT DISULFID 158 179 BY SIMILARITY.
FT AC1_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 37 93 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPPLIC 114 128 PRKOSAEEDLSAEIEN -> CVELSVALLFLFYA (IN ISOFORM 3).
FT VARSPPLIC 129 252 MISSING (IN ISOFORM 3).
FT SEQUENCE 252 AA; 27914 MW; B2FF4B86924D4F8F CRC64;

```

```

RP "Sequence of a full length cDNA coding for human protein HC (alpha 1
RT microglobulin).";
RT Nucleic Acids Res. 14:6340-6340(1986).
RN [5]
RP SEQUENCE OF 20-202 (INDIVIDUAL WITH TUBULAR PROTEINURIA).
RX MEDLINE=84126849; PubMed=6198962;
RA Lopez C., Grubb A.O., Mendez E.;
RT "The complete amino acid sequence of human complex-forming
RT glycoprotein heterogeneous in charge (protein HC) from one
RT individual.";
RL Arch. Biochem. Biophys. 228:544-554(1984).
RN [6]
RP SEQUENCE OF 20-198 (VARIANT).
RA Lopez C., Grubb A.O., Mendez E.;
RT "Human protein HC displays variability in its carboxyl-terminal amino
RT acid sequence.";
RL FEBS Lett. 144:349-353(1982).
RN [7]
RP SEQUENCE OF 20-198 (PATIENTS WITH TUBULAR PROTEINURIA).
RX MEDLINE=81184038; PubMed=6164372;
RA Takagi T., Takagi K., Kawai T.;
RT "Complete amino acid sequence of human alpha 1-microglobulin.";
RL Biochem. Biophys. Res. Commun. 98:997-1001(1981).
RN [8]
RP SEQUENCE OF 206-350.
RX MEDLINE=85225968; PubMed=2408638;
RA Reisinger P., Hochstrasser K., Albrecht G.J., Lampart K., Saller J.P.;
RT "Human inter-alpha-trypsin inhibitor: localization of the Kunitz-type
RT domains in the N-terminal part of the molecule and their release by a
RT trypsin-like proteinase.";
RL Biol. Chem. Hoppe-Seyler 366:479-483(1985).
RN [9]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=82074265; PubMed=6171497;
RA Hochstrasser K., Schoenberger O.L., Rossmann I., Wächter E.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RT the inter-alpha-trypsin inhibitor, V. Attachments of carbohydrates in
RT the human urinary trypsin inhibitor isolated by affinity
RT chromatography.";
RL Hoppe-Seyler's Z. Physiol. Chem. 362:1357-1362(1981).
RN [10]
RP INHIBITORY SITE.
RX MEDLINE=85225940; PubMed=3890890;
RA Moril M., Travis J.;
RT "The reactive site of human inter-alpha-trypsin inhibitor is in the
RT amino-terminal half of the protein.";
RL Biol. Chem. Hoppe-Seyler 366:19-21(1985).
RN [11]
RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE=90306345; PubMed=1694784;
RA Escribano J., Lopez-Otin C., Hjerpe A., Grubb A., Mendez E.;
RT "Location and characterization of the three carbohydrate prosthetic
RT groups of human protein HC.";
RL FEBS Lett. 266:167-170(1990).
RN [12]
RP CHROMOPHORE.
RC TISSUE=urine;
RX MEDLINE=91340714; PubMed=1714898;
RA Escribano J., Grubb A., Calero M., Mendez E.;
RT "The protein HC chromophore is linked to the cysteine residue at
RT position 34 of the polypeptide chain by a reduction-resistant bond
RT and causes the charge heterogeneity of protein HC.";
RL J. Biol. Chem. 266:15758-15763(1991).
RN [13]
RP SEQUENCE OF 206-219, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
RC TISSUE=Plasma;
RX MEDLINE=94229087; PubMed=7513643;
RA Morelle W., Capon C., Balauduc M., Sautiere P., Kouach M.,
RA Michalski C., Fournet B., Mizon J.;
RT "Chondroitin sulphate covalently cross-links the three polypeptide
RT chains of inter-alpha-trypsin inhibitor.";
RL Eur. J. Biochem. 221:881-888(1994).
RN [14]

```


RP SEQUENCE OF 206-223. AND CROSS-LINK SITE TO HC2.
 RX MEDLINE=93232026; PubMed=7682553;
 RA Enghild J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
 RA Plazo S.V., Hefta S.A.;
 RT "Presence of the protein-glycosaminoglycan-protein covalent cross-link
 in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
 2/dikunin.";
 RT J. Biol. Chem. 268:8711-8716(1993).
 RN [15]
 RP SEQUENCE OF 206-223. AND CROSS-LINK SITE TO HC3.
 RX MEDLINE=91093267; PubMed=1898736;
 RA Enghild J.J., Salvesen G., Hefta S.A., Thøgersen I.B.,
 RA Rutherford S., Plazo S.V.;
 RT "Chondroitin 4-sulfate covalently cross-links the chains of the human
 blood protein pre-alpha-inhibitor.";
 RT J. Biol. Chem. 266:747-751(1991).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 230-339.
 RX MEDLINE=98227321; PubMed=9566199;
 RA Xu Y., Carr P.D., Guss J.M., Ollis D.L.;
 RT "The crystal structure of dikunin from the inter-alpha-inhibitor
 complex: a serine protease inhibitor with two kunitz domains.";
 RL J. Mol. Biol. 276:955-966(1998).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN.
 CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GLANDULOCYTIC
 CC ELASTASE. ADDITIONAL PROTEOLYTIC PROCESSING IN THE KIDNEY AND/OR
 CC URINE CAN PRODUCE FURTHER AMINO- AND CARBOXYL-END MODIFICATIONS
 CC IN ITS SEQUENCE.
 CC -1- SUBUNIT: INTER-ALPHA-TRYPSIN INHIBITOR CONSIST OF A LIGHT CHAIN
 CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
 CC -1- PPM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PPM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
 CC CROSS-LINKING BETWEEN THE DIFFERENT COMPONENTS.
 CC MISCELLANEOUS: IN VITRO, THE FIRST TWELVE RESIDUES OF THE AMINO
 CC END OF THE INHIBITOR APPEAR TO HAVE A REACTIVE SITE CAPABLE OF
 CC INHIBITING THE ACTIVITY OF A NUMBER OF ENZYMES. ITS IN VIVO
 CC FUNCTION IS NOT KNOWN.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54816; CA38585.1; -;
 DR EMBL: X54817; CA38585.1; JOINED.
 DR EMBL: X54818; CA38585.1; JOINED.
 DR EMBL: X04225; CA27803.1; -;
 DR EMBL: M88249; AA559196.1; -;
 DR EMBL: M88165; AA559196.1; JOINED.
 DR EMBL: M88244; AA559196.1; JOINED.
 DR EMBL: M88246; AA559196.1; JOINED.
 DR EMBL: M88247; AA559196.1; JOINED.
 DR EMBL: X04494; CA28182.1; -;
 DR EMBL: X54817; CA38586.1; -;
 DR PIR: A03217; HCHU.
 DR PIR: A25303; A25303.
 DR PIR: S13433; S13433.
 DR PIR: S10717; S10717.
 DR PDB: 1BIK; 16-MAR-99.
 DR SWISS-2DPAGE: P02760; HUMAN.

DR MIM: 176870; -;
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR002345; Lipocalin.
 DR InterPro: IPR000566; Lipocalin_cytfABP.
 DR Pfam: PF00014; Kunitz_BPTI_2.
 DR Pfam: PF00061; Lipocalin_1.
 DR PRINTS: PR00179; LIPOCALIN.
 DR PRINTS: PR00759; BASICPRASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 DR PROSITE: PS00213; LIPOCALIN; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 Query Match 30.1%; Score 94.5; DB 1; Length 352;
 Best Local Similarity 33.3%; Pred. No. 0.0001;
 Matches 18; Conservative 8; Mismatches 23; Indels 5; Gaps 1;
 OY 5 VOACT-----GPTSPHLVLMHYDPQRCGCMTPARGCDGARGFPEYACQOAC 53
 DB 284 VAACMLPIYRGPCRAFIQLMAFDVAKKCVLPFGCGCGNGKFTSEKREYIC 337
 RESULT 3
 ID TFP2_HUMAN STANDARD; PRT; 235 AA.
 AC P48307;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR 2 PRECURSOR (TFPI-2) (PLACENTAL
 DE PROTEIN 5) (Pp5).
 GN TFP12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_Taxid=9606;
 OX [1]
 RN RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP TISSUE-Placenta;
 RC MEDLINE=95204397; PubMed=7896752;
 RA Miyagi Y., Koshikawa N., Yasumitsu H., Miyagi E., Hiraehara F.,
 RA Aoki I., Mitsu K., Umeda M., Miyazaki K.;
 RT "cDNA cloning and mRNA expression of a serine proteinase inhibitor
 RT secreted by cancer cells: identification as placental protein 5 and
 RT tissue factor pathway inhibitor-2.";
 RL J. Biochem. 116:939-942(1994).
 RN [2]
 RN RN SEQUENCE FROM N.A.
 RP TISSUE-Placenta;
 RC MEDLINE=94211862; PubMed=8159751;
 RA Sprecher C.A., Kistel W., Mathewes S., Foster D.C.;
 RT "Molecular cloning, expression, and partial characterization of a
 RT second human tissue-factor-pathway inhibitor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3353-3357(1994).
 RN [3]
 RN RN SEQUENCE FROM N.A.
 RP Maqgi L.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP PARTIAL SEQUENCE OF 23-35; 47-53 AND 133-146.
 RC TISSUE-Placenta;
 RX MEDLINE=88106628; PubMed=3276312;
 RA Buetzow R., Huhtala M.-L., Bohn H., Virtanen I., Seppälä M.;
 RT "Purification and characterization of placental protein 5.";
 RL Biochem. Biophys. Res. Commun. 150:483-490(1988).
 RN [5]
 RP ERRATUM.
 RA Buetzow R., Huhtala M.-L., Bohn H., Virtanen I., Seppälä M.;
 RL Biochem. Biophys. Res. Commun. 151:630-631(1988).
 CC -1- FUNCTION: SEEMS TO INHIBIT TRYPSIN, FACTOR VII(A)/TISSUE FACTOR,
 CC WEAKLY FACTOR XA. HAS NO EFFECT ON THROMBIN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.

OS Mus musculus (Mouse).
OC Fukaryoja, Metzozoa: Chordata: Craniata: Vertebrata; Euteleostomi;
OC Mammalia, Eutheria: Rodentia: Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RC STRAIN-BALB/C; TISSUE-Brain;
RX MEDLINE=92066458; PubMed=1756177;
RA de Strooper B., van Leuven F., van den Berghe H.;
RT "The amyloid beta protein precursor or proteinase nexin II from mouse
RL is closer related to its human homolog than previously reported.";
RN Biochim. Biophys. Acta 1129:141-143(1991).
RN (2)
RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=88106489; PubMed=3322280;
RA Yamada T., Sasaki H., Furuya H., Miyata T., Goto I., Sasaki Y.;
RT "Complementary DNA for the mouse homolog of the human amyloid beta
RL protein precursor.";
RN Biochem. Biophys. Res. Commun. 149:665-671(1987).
RN (3)
RP REVISIONS.
RA Yamada T.;
RL Submitted (MAR-1988) to the EMBL/Genbank/DBJ databases.
RN (4)
RP SEQUENCE OF 289-364 FROM N.A.
RC STRAIN=CD-1; TISSUE=Placenta;
RX MEDLINE=89345111; PubMed=2569710;
RA Fukuchi K., Martin G.M., Deeb S.S.;
RT "Sequence of the protease inhibitor domain of the A4 amyloid protein
RL precursor of Mus domesticus.";
RN Nucleic Acids Res. 17:5396-5396(1989).
RN (5)
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE=92209998; PubMed=1555768;
RA Izumi R., Yamada T., Yoshikai S.I., Sasaki H., Hattori M.,
RA Sakai Y.;
RT "Positive and negative regulatory elements for the expression of the
RL Alzheimer's disease amyloid precursor-encoding gene in mouse.";
RN Gene 112:189-195(1992).
RN (6)
RP SEQUENCE OF 281-380 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain and Kidney;
RX MEDLINE=89149613; PubMed=2493250;
RA Yamada T., Sasaki H., Dohura K., Goto I., Sakaki Y.;
RT "Structure and expression of the alternatively-spliced forms of mRNA
RL for the mouse homolog of Alzheimer's disease amyloid beta protein
RT precursor.";
RN Biochem. Biophys. Res. Commun. 158:906-912(1989).
RN (7)
RP SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
RN (8)
RP ALTERNATIVE PRODUCTS: 5 ISOFORMS: APP(395), APP(563), APP(695),
RX APP(751) AND APP(770) (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE
RN SPLICING.
RN (9)
RP TISSUE SPECIFICITY: AAA(770) IS EXPRESSED IN KIDNEY. AAA(751) IS
RX WIDELY EXPRESSED. AAA(695) IS EXPRESSED IN BRAIN, KIDNEY AND
RN LIVER.
RN (10)
RP DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
RX WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
RN RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
RX NEXT MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
RN PHOSPHORYLATION (BY SIMILARITY).
RN (11)
RP SIMILARITY: BELONGS TO THE APP FAMILY.
RN (12)
RP SIMILARITY: CONTAINS 1 BPRT/WUNTZ INHIBITOR DOMAIN.
RN (13)
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RX between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RX use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RX entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN (14)
RX EMBL; X59379.; NOT_ANNOTATED_CDS

DR EMBL: M18373; AAA37139.1; -
 DR EMBL: X15210; CAA33280.1; -
 DR EMBL: D10603; BAA01456.1; -
 DR EMBL: M24397; AAA39929.1; -
 DR PIR: A27485; A27485.
 DR PIR: S04855; S04855.
 DR PIR: S19727; S19727.
 DR MGI: 88059; APP.
 DR InterPro: IPR001868; A4_APP.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF02177; A4_EXTRA; 1.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR PRINTS: PR00203; AMYLOID4.
 DR PRINTS: PR00204; BETAAMYLOID.
 DR PRINTS: PR00759; BASICTRIASE.
 DR SMART: SM00006; A4_EXTRA; 1.
 DR SMART: SM00131; KU; 1.
 DR PROSITE: PS00319; A4_EXTRA; 1.
 DR PROSITE: PS00320; A4_INTRA; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 1.
 DR Glycoprotein: Amyloid; Neurone; Transmembrane; Signal;
 KW Alternative splicing; Serine protease inhibitor.
 FT SIGNAL 1 17
 FT CHAIN 18 770
 FT DOMAIN 18 699
 FT TRANSMEM 700 723
 FT DOMAIN 724 770
 FT DOMAIN 673 715
 FT DOMAIN 287 345
 FT SITE 759 762
 FT DISULFID 281 341
 FT DISULFID 300 324
 FT DISULFID 316 337
 FT CARBOHD 542 542
 FT CARBOHD 571 571
 FT VARSPPLIC 289 364
 FT VARSPPLIC 346 380
 FT SEQUENCE 770 AA; 86752 MW; 26C50DE0890CA7A CRC64;
 SQ
 Query Match 29.9%; Score 94; DB 1; Length 770;
 Best Local Similarity 37.5%; Pred. No. 0.00025;
 Matches 18; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

6 QACTGPTSPHLVLMHYDPQRCGCMTPFARGCDGAARGFETVEACQAC 53
 DB 294 QAGTGPCRAMISRWFEDVTEGKCVFFYGGCGGNRNFTETECMAVC 341

RESULT 5
 A4_MACMU STANDARD: PRT: 76 AA.
 AC P29216:
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN (FRAGMENT).
 GN APP.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OC NCBI_TaxID=9544;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Koo E.H., Sisodia S.S., Price D.L.;
 RA Submitted (JUL-1989) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
 CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN

CC G(O) (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 5 ISOFORMS; APP(353), APP(563), APP(635),
 CC APP(751) AND APP(770) (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE
 CC SPLICING (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: X15985; CAA34116.1; -
 CC PIR: S06678; S06678.
 CC HSP: P05067; ICAO.
 DR InterPro: IPR001868; A4_APP.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR SMART: SM00131; KU; 1.
 DR PROSITE: PS00319; A4_EXTRA; PARTIAL.
 DR PROSITE: PS00320; A4_INTRA; PARTIAL.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 1.
 DR Glycoprotein: Amyloid; Neurone; Alternative splicing;
 KW Serine protease inhibitor.
 FT NON_TER 1 1
 FT DOMAIN 1 76
 FT ACT_SITE 13 14
 FT DISULFID 3 53
 FT DISULFID 12 36
 FT DISULFID 28 49
 FT NON_TER 76 76
 FT SEQUENCE 76 AA; 8527 MW; 492BF3069AB082A1 CRC64;
 SQ
 Query Match 29.3%; Score 92; DB 1; Length 76;
 Best Local Similarity 37.5%; Pred. No. 4.6e-05;
 Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

6 QACTGPTSPHLVLMHYDPQRCGCMTPFARGCDGAARGFETVEACQAC 53
 DB 6 QAGTGPCRAMISRWFEDVTEGKCVFFYGGCGGNRNFTETECMAVC 33

RESULT 6
 A4_MACFA STANDARD: PRT: 87 AA.
 AC P53601:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN (FRAGMENT).
 GN APP.
 OS Macaca fascicularis (Craab eating macaque) (Cynomolous monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OC NCBI_TaxID=9541;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91273117; PubMed=1905108;
 RA Podlinsky M.B., Tolan D.R., Selkoe D.J.;
 RT "Homology of the amyloid beta protein precursor in monkey and human
 RT supports a primate model for beta amyloidosis in Alzheimer's
 RT disease";
 RL Am. J. Pathol. 138:1423-1435(1991).
 CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
 CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
 CC G(O) (BY SIMILARITY).

```

CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
CC PHOSPHORYLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC
CC EMBL: S81024; AADI4347.1; -.
CC InterPro: IPR001868; A4_APP.
CC InterPro: IPR002223; Kunitz_BPTI.
CC Pfam: PF02177; A4_EXTRA: 1.
CC Pfam: PF00014; Kunitz_BPTI: 1.
CC PRINTS: PR00203; AMTLOID4.
CC PRINTS: PR00204; BETAMYLLOID.
CC PRINTS: PR00759; BASICPTASE.
CC SMART: SM00006; A4_EXTRA: 1.
CC SMART: SM00131; KU: 1.
CC PROSITE: PS00319; A4_EXTRA: 1.
CC PROSITE: PS00320; A4_IMPA: 1.
CC PROSITE: PS00280; BPTI_KUNITZ_1; 1.
CC PROSITE: PS50279; BPTI_KUNITZ_2; 1.
CC GlycoProtein: Amyloid; Neurone; Transmembrane; Alternative splicing;
CC Signal; Serine protease inhibitor.
CC KW SIGNAL.
CC FT CHAIN 1 17 BY SIMILARITY.
CC FT CHAIN 18 751 A4 PROTEIN.
CC FT CHAIN 18 680 BETA-AMYLOID PROTEIN (POTENTIAL).
CC FT TRANSMEM 681 704 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 705 751 POTENTIAL.
CC FT DOMAIN 287 345 CYTOPLASMIC (POTENTIAL).
CC FT SITE 301 743 BPTI/KUNITZ INHIBITOR.
CC FT ACT_SITE 301 302 CLATHRIN-BINDING (BY SIMILARITY).
CC FT DISULFID 291 341 REACTIVE BOND.
CC FT DISULFID 300 324 BY SIMILARITY.
CC FT DISULFID 316 337 BY SIMILARITY.
CC FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (PROBABLE).
CC FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (PROBABLE).
CC SQ SEQUENCE 751 AA; 84893 MW; 6C3F431089569049 CRC64;

Query Match 29.38; Score 92; DB 1; Length 751;
Best Local Similarity 37.58; Pred. No. 0.00044;
Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 6 QACTGPTSPHLVLYHYDPQRCGCMTPFARGCDGAARGFETYEAGQAG 53
  ||||| :||:| :||| :||| :||| :||| :||| :||| :||| :|||
Db 294 QAEIGPCRAMISRWYFDYTBCKCAPFYGGCGGRNNFDEBYCYMAVC 341

RESULT 8
APP2_HUMAN
ID APP2_HUMAN STANDARD; PRT; 763 AA.
AC 006481;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AMYLOID-LIKE PROTEIN 2 PRECURSOR (AMYLOID PROTEIN HOMOLOG) (APPH)
DE (CDEI-BOX BINDING PROTEIN) (CDEBP).
GN APP2 OR APP12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
SEQUENCE FROM N.A.

```


DR SMART; SM00006; A4-EXTRA; 1.
 DR SMART; SM00131; KU; 1.
 DR PROSITE; PS00319; A4-EXTRA; 1.
 DR PROSITE; PS00320; A4-EXTRA; 1.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
 DR Transmembrane; Alternative splicing; Serine protease inhibitor;
 KW Signal; Glycoprotein.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 765 AMYLOID-LIKE PROTEIN 2.
 FT DOMAIN 30 695 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 696 718 POTENTIAL.
 FT DOMAIN 719 765 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 218 282 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT ACT_SITE 322 323 BPTI/KUNITZ INHIBITOR.
 FT DISULFID 312 362 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 321 345 BY SIMILARITY.
 FT DISULFID 337 358 BY SIMILARITY.
 FT DOMAIN 218 229 POLY-GLU.
 FT CARBOHYD 628 628 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
 FT VARSPLIC 311 365 MISSING (IN ISOFORM C AND ISOFORM D).
 FT VARSPLIC 616 627 MISSING (IN ISOFORM B AND ISOFORM D).
 FT CONFLICT 575 577 DOF -> EPV (IN REF. 2).
 SO SEQUENCE 765 AA; 86682 MW; CP51PCCE305AOCF CRC64;

Query Match 29.3%; Score 92; DB 1; Length 765;
 Best Local Similarity 33.3%; Pred. No. 0.00045;
 Matches 16; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

OY 6 OACGPTSPHVLWHDYDQGGCTFPARCGDGAARCFETEAQOQAC 53
 Db 315 EAMTGPCRAVPRWTFDLKSGCVFTYGGCGNNRNNSESDYCMAYC 362

RESULT 10
 A4_HUMAN
 ID A4_HUMAN STANDARD; PRT; 770 AA.
 AC P05067; P09000; Q16011;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN PRECURSOR (PROTEASE NEXIN-II)
 DE (PNI-II) (ABPI) [CONTAINS: BETA-AMYLOID PROTEIN (BETA-APP) (A-BETA)].
 GN APP OR A4 OR CAPP OR ADL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=87144572; PubMed=2881207;
 RA Kang J., Lemaitre H.-G., Unterbeck A., Salpaum J.M., Masters C.L.,
 RA Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;
 RT "The precursor of Alzheimer's disease amyloid A4 protein resembles a
 RT cell-surface receptor.";
 RL Nature 325:733-736(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88122639; PubMed=2893289;
 RA Ponte P., Gonzalez-Dewhitt P., Schilling J., Miller J., Hsu D.,
 RA Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F.,
 RA Cordell B.;
 RT "A new A4 amyloid mRNA contains a domain homologous to serine
 RT proteinase inhibitors.";
 RL Nature 331:525-527(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89128427; PubMed=2783775;
 RA Lemaitre H.-G., Salpaum J.M., Multhaup G., Kang J., Bayney R.M.,
 RA Unterbeck A., Beyreuther K., Mueller-Hill B.;

RT "The Pre4(695) precursor protein of Alzheimer's disease A4 amyloid
 RT is encoded by 16 exons.";
 RL Nucleic Acids Res. 17:517-522(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97263807; PubMed=9108164;
 RA Hattori M., Tsukahara F., Furuhata Y., Tanahashi H., Hirose M.,
 RA Saito M., Tsukuni S., Sakaki Y.;
 RT "A novel method for making nested deletions and its application for
 RT sequencing of a 300 kb region of human APP locus.";
 RL Nucleic Acids Res. 25:1802-1808(1997).
 RN [5]
 RP SEQUENCE OF 286-345 AND 365-366 FROM N.A.
 RX MEDLINE=88122640; PubMed=2893290;
 RA Tanzi R.E., McClatchey A.I., Lamperti E.D., Villa-Komaroff L.,
 RA Gusella J.F., Neve R.L.;
 RT "Protease inhibitor domain encoded by an amyloid protein precursor
 RT mRNA associated with Alzheimer's disease.";
 RL Nature 331:528-530(1988).
 RN [6]
 RP SEQUENCE OF 287-367 FROM N.A.
 RX MEDLINE=88122641; PubMed=2893291;
 RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;
 RT "Novel precursor of Alzheimer's disease amyloid protein shows
 RT protease inhibitory activity.";
 RL Nature 331:530-532(1988).
 RN [7]
 RP SEQUENCE OF 284-289 AND 365-770 FROM N.A.
 RX MEDLINE=87231971; PubMed=3035574;
 RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;
 RT "Molecular cloning and characterization of a cDNA encoding the
 RT cerebrovascular and the neuritic plaque amyloid peptides.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).
 RN [8]
 RP SEQUENCE OF 507-770 FROM N.A.
 RX MEDLINE=88124954; PubMed=2893379;
 RA Zain S.B., Salim M., Chou W.G., Sajdel-Sulkowska E.M., Majocha R.E.,
 RA Marotta C.A.;
 RT "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer
 RT disease brain: coding and noncoding regions of the fetal precursor
 RT mRNA are expressed in the cortex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).
 RN [9]
 RP SEQUENCE OF 672-681.
 RX MEDLINE=88035004; PubMed=3312495;
 RA Pardridge W.M., Vinters H.V., Yang J., Eisenberg J., Choi T.B.,
 RA Tourtellotte W.W., Huebner V., Shively J.E.;
 RT "Amyloid angiopathy of Alzheimer's disease: amino acid composition
 RT and partial sequence of a 4,200-dalton peptide isolated from cortical
 RT microvessels.";
 RL J. Neurochem. 49:1394-1401(1987).
 RN [10]
 RP SEQUENCE OF 739-770 FROM N.A.
 RX MEDLINE=90236318; PubMed=2110105;
 RA Yoshikai S.-I., Sasaki H., Don-Ura K., Furuya H., Sakaki Y.;
 RT "Genomic organization of the human amyloid beta-protein precursor
 RT gene.";
 RL Gene 87:257-263(1990).
 RN [11]
 RP SEQUENCE OF 1-10 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=89016647; PubMed=3140222;
 RA Schon E.A., Mita S., Sadlock J., Herbert J.;
 RT "A cDNA specifying the human amyloid beta precursor protein (ABPP)
 RT encodes a 95-kDa polypeptide.";
 RL Nucleic Acids Res. 16:9351-9351(1988).
 RN [12]
 RP SEQUENCE OF 18-50.
 RX MEDLINE=87250462; PubMed=3597385;
 RA van Nostrand W.E., Cunningham D.D.;
 RT "Purification of protease nexin II from human fibroblasts.";
 RL J. Biol. Chem. 262:8508-8514(1987).
 RN [13]

RP IDENTITY OF APP WITH NEXIN-II.
 RX MEDLINE-89384866; PubMed-2506449;
 RA Oltersdorf T., Fritz L.C., Schenk D.B., Lieberburg I.,
 RA Johnson-Wood K.L., Beattie E.C., Ward P.J., Blacher R.W., Dovey H.F.,
 RA Slnha S.;
 RT "The secreted form of the Alzheimer's amyloid precursor protein with
 RT the Kunitz domain is protease nexin-II.";
 RL Nature 341:144-147(1989).
 RN [14]
 RP PROTEASE-SPECIFICITY OF INHIBITOR DOMAIN.
 RX MEDLINE-90211252; PubMed-1969731;
 RA Kido H., Fukutomi A., Schilling J., Wang Y., Cordell B., Katunuma N.;
 RT "Protease-specificity of Kunitz inhibitor domain of Alzheimer's
 RT disease amyloid protein precursor.";
 RL Biochem. Biophys. Res. Commun. 167:716-721(1990).
 RN [15]
 RP COMPLEX WITH G(O).
 RX MEDLINE-93188965; PubMed-8446172;
 RA Nishimoto I., Okamoto T., Matsura Y., Takahashi S., Okamoto T.,
 RA Murayama Y., Ogata E.;
 RT "Alzheimer amyloid protein precursor complexes with brain GTP-binding
 RT protein G(O).";
 RL Nature 362:75-79(1993).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 28-133.
 RX MEDLINE-99215582; PubMed-10201399;
 RA Rossjohn J., Cappai R., Fell S.C., Henry A., McKinsty W.J.,
 RA Galatis D., Hesse L., Multhaup G., Beyreuther K., Masters C.L.,
 RA Parker M.W.;
 RT "Crystal structure of the N-terminal, growth factor-like domain of
 RT Alzheimer amyloid precursor protein.";
 RL Nat. Struct. Biol. 6:327-331(1999).
 RN [17]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 287-344.
 RX MEDLINE-91104913; PubMed-2125487;
 RA Hynes T.R., Randal M., Kennedy L.A., Eigenbrodt C., Kossiakof A.A.;
 RT "X-ray crystal structure of the protease inhibitor domain of
 RT Alzheimer's amyloid beta-protein precursor.";
 RL Biochemistry 29:10018-10022(1990).
 RN [18]
 RP STRUCTURE BY NMR OF 289-344.
 RX MEDLINE-92031488; PubMed-1718421;
 RA Heald M.L., Tilton R.F. Jr., Hammond L.S., Lee A., Bayney R.M.,
 RA Kamarc M.E., Ramabhadran T.V., Dreyer R.N., Davis G., Unterebeck A.,
 RA Tamburini P.P.;
 RT "Sequential NMR resonance assignment and structure determination of
 RT the Kunitz-type inhibitor domain of the Alzheimer's beta-amyloid
 RT precursor protein.";
 RL Biochemistry 30:10467-10478(1991).
 RN [19]
 RP STRUCTURE BY NMR OF 672-699.
 RX MEDLINE-94281210; PubMed-7516706;
 RA Talafous J., Marciniowski K.J., Klopman G., Zagorski M.G.;
 RT "Solution structure of residues 1-28 of the amyloid beta-peptide.";
 RL Biochemistry 33:7788-7796(1994).
 RN [20]
 RP STRUCTURE BY NMR OF 696-706.
 RX MEDLINE-97128622; PubMed-8973180;
 RA Kohno T., Kobayashi K., Maeda T., Sato K., Takashima A.;
 RT "Three-dimensional structures of the amyloid beta peptide (25-35) in
 RT membrane-mimicking environment.";
 RL Biochemistry 35:16094-16104(1996).
 RN [21]
 RP STRUCTURE BY NMR OF 672-711.
 RX MEDLINE-98359783; PubMed-9693002;
 RA Coles M., Bichenell W., Watson A.A., Fairlie D.P., Craik D.J.;
 RT "Solution structure of amyloid beta-peptide(1-40) in a water-miscelle
 RT environment. Is the membrane-spanning domain where we think it is?";
 RL Biochemistry 37:11064-11077(1998).
 RN [22]
 RP STRUCTURE BY NMR OF 672-699.
 RX MEDLINE-20400066; PubMed-10940222;
 RA Poulsen S.-A., Watson A.A., Craik D.J.;

RT "Solution structures in aqueous SDS micelles of two amyloid beta
 RT peptides of Abeta(1-28) mutated at the alpha-secretase cleavage
 RT site.";
 RL J. Struct. Biol. 130:142-152(2000).
 RN [23]
 RP SIGNAL SEQUENCE CLEAVAGE SITE, AND TOPOLOGY.
 RX MEDLINE-88296437; PubMed-2900137;
 RA Dykes T., Weidemann A., Multhaup G., Salbaum J.M., Lemaitre H.-G.,
 RA Kang J., Mueller-Hill B., Masters C.L., Beyreuther K.;
 RT "Identification, transmembrane orientation and biogenesis of the
 RT amyloid A4 precursor of Alzheimer's disease.";
 RL EMBO J. 7:949-957(1988).
 RN [24]
 RP REVIEW.
 RX MEDLINE-92271194; PubMed-1589757;
 RA Kosik K.S.;
 RT "Alzheimer's disease: a cell biological perspective.";
 RL Science 256:780-783(1992).
 RN [25]
 QY 6 OACTGPTSPHLVLMHYDPQGGCMTPPARGCDGAGFETYEACQAC 53
 DB 294 QAGTGCAMISRMVFDYEGKCAFFYGGCGGRNNRNFDEYCMAYC 341
 Query Match 29.3%; Score 92; DB 1; Length 770;
 Best Local Similarity 37.5%; Pred. No. 0.00045;
 Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;
 RESULT 11
 ID A4_RAT STANDARD: PRT: 770 AA.
 AC P08592:
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALZHEIMER'S DISEASE AMYLOID A4 PROTEIN HOMOLOG PRECURSOR
 DE (AMYLOIDOGENIC GLYCOPROTEIN) (AG).
 GN APP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE-88312583; PubMed-2900758;
 RA Shivers B.D., Hilbich C., Multhaup G., Salbaum J.M., Beyreuther K.,
 RA Seeburg P.H.;
 RT "Alzheimer's disease amyloidogenic glycoprotein: expression pattern
 RT in rat brain suggests a role in cell contact.";
 RL EMBO J. 7:1365-1370(1988).
 RN [2]
 RP SEQUENCE OF 289-364 FROM N.A.
 RC TISSUE=liver;
 RX MEDLINE-89183625; PubMed-2648331;
 RA Kang J., Mueller-Hill B.;
 RT "The sequence of the two extra exons in rat preA4.";
 RL Nucleic Acids Res. 17:2130-2130(1989).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS: APP(395), APP(563), APP(695),
 CC APP(751) AND APP(770) (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -1- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
 CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
 CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
 CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
 CC PHOSPHORYLATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BP1/KONIT2 INHIBITOR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss institute of Bioinformatics and the EMBL outstation -


```

RT      inter-alpha-trypsin inhibitor ";
RL      Biol. Chem. Hoppe-Seyler 368:727-731(1987).
RN      [2]
RP      SEQUENCE.
RC      SPECIES-C. hircus;
RA      MEDLINE=90105540; PubMed=2481505;
RX      Rasm G., Hochstrasser K., Gerl C., Machter E.;
RT      "Primary structure of a proteinase inhibitor released from goat serum
RT      inter-alpha-trypsin inhibitor.";
RL      Biochim. Biophys. Acta 999:335-337(1989).
CC      -1- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE IT1 AFTER
CC      LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS.
CC      WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE
CC      FIRST DOMAIN INTERACTS WEAKLY WITH PAM-GRANULOCYTTIC ELASTASE AND
CC      NOT AT ALL WITH PANCREATIC ELASTASE.
CC      -1- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO
CC      DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN 1.
CC      INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH
CC      CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC      -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
DR      PIR: A29653; A29652.
DR      HSSP: P10646; IAD2.
DR      InterPro: IPR002223; Kunitz_BPTI.
DR      Pfam: PF00014; Kunitz_BPTI; 2.
DR      SMART: SM00131; KU; 2.
DR      PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR      PROSITE: PS0279; BPTI_KUNITZ_2; 2.
KW      Plasma; Glycoprotein; Serine protease inhibitor; repeat.
FT      NON_TER      1      1
FT      DOMAIN      5      55      BPTI/KUNITZ INHIBITOR 1.
FT      DISULFID      61      111      BPTI/KUNITZ INHIBITOR 2.
FT      DISULFID      5      55      BY SIMILARITY.
FT      DISULFID      14      38      BY SIMILARITY.
FT      DISULFID      30      51      BY SIMILARITY.
FT      DISULFID      61      111      BY SIMILARITY.
FT      DISULFID      70      94      BY SIMILARITY.
FT      DISULFID      86      107      BY SIMILARITY.
FT      ACT_SITE      15      16      INHIBITORY SITE (P1) (CHYMOTRYPSIN,
FT      ELASTASE).
FT      ACT_SITE      71      72      INHIBITORY SITE (P1) (TRYPSIN).
FT      CARBOHD      24      24      N-LINKED (GLCNAC. . .).
FT      NON_TER      123
SQ      SEQUENCE      123 AA; 13686 MW; 295038173F22D2D1 CRC64;

Query Match      29.1%; Score 91.5; DB 1; Length 123;
Best Local Similarity      33.3%; Pred. No. 8.5e-05;
Matches      18; Conservative      9; Mismatches      22; Indels      5; Gaps      1;

Oy      5 VOACT-----GFTSPHLVLMHYDPOFGCGMTFPAFGCDGARGFTFYERACQAC      53
      |||||      |||      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
Db      58 VOACMLPIVRGRCRAGIELMAFDVAVKRCVREIFYGCGNGNGNQFYSOKEKEYC      111

RESULT      13
AMBIP_BOVIN
ID      AMBP_BOVIN      STANDARD;      PRT;      352 AA.
AC      P00978; P35420; Q28020;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE      TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30) (BI-14)
DE      (CMOLUS EXTRACELLULAR MATRIX STABILIZING FACTOR) (ESF)].
DE      AMP OR ITII.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheta; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
      [1]
SEQUENCE FROM N.A.
PP      TISSUE-layer:

```


RX MEDLINE=96201710; PubMed=8611630;
RA Lindqvist A., Akerstrom B.;
RT "Bovine alpha 1-microglobulin/dikunin. Isolation and characterization
of liver cDNA and urinary alpha 1-microglobulin.";
RL Biochim. Biophys. Acta 1306:98-106(1996).
RN (2)
RP SEQUENCE OF 227-349.
RX MEDLINE=85225967; PubMed=2408637;
RA Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
the inter-alpha-trypsin inhibitor, X. The amino-acid sequences of the
trypsin-released inhibitors from horse and pig inter-alpha-trypsin
inhibitors.";
RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).
RN [3]
RP SEQUENCE OF 227-348.
RX MEDLINE=84133807; PubMed=6199275;
RA Hochstrasser K., Wachter E.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
the inter-alpha-trypsin inhibitor, VII. Determination of the
amino-acid sequence of the trypsin-released inhibitor from bovine
inter-alpha-trypsin inhibitor.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1679-1687(1983).
RN [4]
RP SEQUENCE OF 206-219.
RX TISSUE=Fetal serum;
RA MEDLINE=92291130; PubMed=1376324;
RA Chen L., Mao S.J.T., Larsen W.J.;
RT "Identification of a factor in fetal bovine serum that stabilizes the
cumulus extracellular matrix. A role for a member of the inter-alpha-
trypsin inhibitor family.";
RL J Biol. Chem. 267:12380-12386(1992).
RN [5]
RP REACTIVE SITES.
RX MEDLINE=84133808; PubMed=6199276;
RA Hochstrasser K., Albrecht G.J., Schoenberger O.L., Wachter E.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
the inter-alpha-trypsin inhibitor, VII. Characterization of the
bovine inhibitor as double-headed trypsin-elastase inhibitor.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1689-1696(1983).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN.
CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTTIC
CC ELASTASE.
CC -1- FUNCTION: MAY DIFFUSE INTO FOLLICULAR FLUID AFTER AN OVULATORY
CC STIMULUS TO ACT AS STRUCTURAL LINKER THAT ENSURE NORMAL CUMULUS
CC EXPANSION, THROUGH STABILIZATION OF THE CUMULUS EXTRACELLULAR
CC MATRIX THUS SUPPORTING THE PROCESS OF OVULATION.
CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U35642; AAB07599.1; -;
DR PIR: A01209; TIBOBI.
DR HSP: P10646; 1NDZ.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR002345; Lipocalin.
DR InterPro: IPR000566; Lipocalin_cytfabp.
DR Pfam: PF00014; Kunitz_BPTI; 2.

DR Pfam: PF00061; Lipocalin; 1.
DR PRINTS: PR00179; LIPOCALIN.
DR PRINTS: PR00759; BASICPHASE.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 2.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
KW Lipocalin.
FT SIGNAL 1 19
FT CHAIN 20 203
FT CHAIN 206 352
FT
FT DOMAIN 231 281
FT BINDING 287 337
FT BINDING 53 53
FT DISULFID 91 188
FT DISULFID 231 281
FT DISULFID 240 264
FT DISULFID 256 277
FT DISULFID 287 337
FT DISULFID 296 320
FT DISULFID 312 333
FT ACT_SITE 241 242
FT
FT ACT_SITE 297 298
FT CARBOHD 115 115
FT CARBOHD 223 223
FT CARBOHD 250 250
FT CONFLICT 209 209
FT CONFLICT 217 217
FT CONFLICT 268 268
FT CONFLICT 274 274
FT CONFLICT 298 299
FT CONFLICT 330 330
FT CONFLICT 346 346
FT SEQUENCE 352 AA; 39235 MM; ED1CGCA02E70B19 CnC64;

Query Match 29.1%; Score 91.5; DB 1; Length 352;
Best Local Similarity 31.5%; Pred. No. 0.0024;
Matches 17; Conservative 10; Mismatches 22; Indels 5; Gaps 1;

OY 5 VQACT-----GPTSPHIVMHDPRGCGMTTPAGCCGSAAGFEYEGCGAC 53
DB 284 VECMLPIVGGPCRSYIQLMADFADKVCVRPSGGCKGNKFESECKEYCG 337

RESULT 14
ID AMP_MERUN STANDARD; PRT; 346 AA.
AC 062577; 062576;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPsin INHIBITOR LIGHT CHAIN (ITI-1C) (BIKUNIN) (HI-30)].
GN AMP OR ITIL.
OS Meriones unguiculatus (Mongolian jird).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Meriones.
OX NCBI_TaxID=10047;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95110820; PubMed=7529051;
RA Ide H., Itoh H., Nawa Y.;
RT "Sequencing of cDNAs encoding alpha 1-microglobulin/dikunin of
RT Mongolian gerbil and Syrian golden hamster in comparison with man and
RT other species.";
RL Biochim. Biophys. Acta 1209:286-292(1994).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL

CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN (BY SIMILARITY).
 CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE (BY SIMILARITY).
 CC -1- PFM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PFM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D31813; BAA06600.1; -
 CC HSP: P10646; 1ND2.
 CC InterPro: IPR002223; Kunitz_BPTI.
 CC InterPro: IPR002345; Lipocalin.
 CC InterPro: IPR000566; Lipocalin_cytfabp.
 CC Pfam: PF00014; Kunitz_BPTI; 2.
 CC Pfam: PF00061; Lipocalin; 1.
 CC PRINTS: PR00179; LIPOCALIN.
 CC PRINTS: PR00759; BASICPTASE.
 CC SMART: SM00131; KU; 2.
 CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 CC PROSITE: PS00279; BPTI_KUNITZ_2; 2.
 CC Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 CC Lipocalin.
 CC FT SIGNAL 1 19 BY SIMILARITY.
 CC FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
 CC FT CHAIN 205 346 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
 CC CHAIN.
 CC DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 CC DOMAIN 286 336 BPTI/KUNITZ INHIBITOR 2.
 CC BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 CC DISULFID 90 187 BY SIMILARITY.
 CC DISULFID 230 280 BY SIMILARITY.
 CC DISULFID 239 263 BY SIMILARITY.
 CC DISULFID 255 276 BY SIMILARITY.
 CC DISULFID 286 336 BY SIMILARITY.
 CC DISULFID 295 319 BY SIMILARITY.
 CC DISULFID 311 332 BY SIMILARITY.
 CC CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC ACT_SITE 240 241 INHIBITOR SITE (PI) (CHYMOTRYPSIN,
 CC ELASTASE) (BY SIMILARITY).
 CC ACT_SITE 296 297 INHIBITOR SITE (PI) (TRYPSIN) (BY
 CC SIMILARITY).
 CC SEQUENCE 346 AA; 38643 MW; F1A463810918D5F CRC64;

Query Match 28.8%; Score 90.5; DB 1; Length 346;
 Best Local Similarity 34.6%; Pred. No. 0.00032;
 Matches 18; Conservative 9; Mismatches 22; Indels 3; Gaps 1;

OY 2 LPVQACTGPTSPHLVIMHYDPQGGCMTPARGCDGANGFETYEACQAC 53
 Db 288 LPVQACTGPTSPHLVIMHYDPQGGCMTPARGCDGANGFETYEACQAC 53

RESULT 15
 AMP_MESAU STANDARD; PRT; 349 AA.
 ID AMP_MESAU
 AC Q60559; Q60558;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
 GN AMP OR ITIL.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_Taxid=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95110820; PubMed=7529051;
 RA Ide H., Itoh H., Nawa Y.;
 RT "Sequencing of cDNAs encoding alpha 1-microglobulin/bikunin of
 RT Mongolian gerbil and Syrian golden hamster in comparison with man and
 RT other species";
 RL Biochim. Biophys. Acta 1209:286-292(1994).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN (BY SIMILARITY).
 CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE (BY SIMILARITY).
 CC -1- PFM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PFM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D31814; BAA06601.1; -
 CC HSP: P10646; ITPX.
 CC InterPro: IPR002223; Kunitz_BPTI.
 CC InterPro: IPR002345; Lipocalin.
 CC InterPro: IPR000566; Lipocalin_cytfabp.
 CC Pfam: PF00014; Kunitz_BPTI; 2.
 CC Pfam: PF00061; Lipocalin; 1.
 CC PRINTS: PR00179; LIPOCALIN.
 CC PRINTS: PR00759; BASICPTASE.
 CC SMART: SM00131; KU; 2.
 CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 CC PROSITE: PS00279; BPTI_KUNITZ_2; 2.
 CC Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 CC Lipocalin.
 CC FT SIGNAL 1 19 BY SIMILARITY.
 CC FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
 CC FT CHAIN 205 349 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
 CC CHAIN.
 CC DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 CC DOMAIN 286 336 BPTI/KUNITZ INHIBITOR 2.
 CC BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 CC DISULFID 90 187 BY SIMILARITY.
 CC DISULFID 230 280 BY SIMILARITY.
 CC DISULFID 239 263 BY SIMILARITY.
 CC DISULFID 255 276 BY SIMILARITY.
 CC DISULFID 286 336 BY SIMILARITY.
 CC DISULFID 295 319 BY SIMILARITY.
 CC DISULFID 311 332 BY SIMILARITY.
 CC CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT ACT_SITE 240 241 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
 FT ACT_SITE 296 297 ELASTASE) (BY SIMILARITY).
 FT INHIBITORY SITE (P1) (TRYPsin) (BY
 SO SEQUENCE 349 AA; 38782 MW; 8C954584B7DBE728 CRC64;
 SIMILARITY).

Query Match 28.5%; Score 89.5; DB 1; Length 349;
 Best Local Similarity 34.6%; Pred. No. 0.00043;
 Matches 18; Conservative 9; Mismatches 22; Indels 3; Gaps 1;

OY 2 LPDVACTGPTSPHLVLMHYDPORGCMTPARGCDGARGFETYEACQAC 53
 Db 288 LPIVQ---GPCRAYVELMAFDAAAGCKVQFSYGCKNGNKFYSKECKEYEC 336

Search completed: February 26, 2002, 01:46:41
 Job time: 1185 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:29:38 ; Search time 78.07 Seconds
(without alignments)
51.713 Million cell updates/sec

Title: us-09-819-136-2_copy_299_351

Perfect score: 314

Sequence: 1 CLPDVQACTGPTSPHLYLVLMH.....RGCDGARGFETVACQAC 53

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

1: PIR68:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.5	36.5	1743	2 T26859	hypothetical prote
2	104	33.1	252	2 JG0185	hepatocyte growth
3	94.5	30.1	335	2 T32657	hypothetical prote
4	94.5	30.1	352	1 HCHU	alpha-1-microglobu
5	94	29.9	76	2 S04855	Alzheimer's diseas
6	94	29.9	100	2 A32282	Alzheimer's diseas
7	94	29.9	235	2 A34951	tissue factor path
8	92.5	29.5	922	2 T23573	hypothetical prote
9	92	29.3	76	2 S03607	Alzheimer's diseas
10	92	29.3	76	2 S06678	Alzheimer's diseas
11	92	29.3	111	2 S41082	amyloid precursor
12	92	29.3	484	4 A33761	hypothetical Alzhe
13	92	29.3	751	2 A49974	beta-amyloid precu
14	92	29.3	763	2 A49321	amyloid beta (A4)
15	92	29.3	765	2 S42880	amyloid precursor-
16	92	29.3	770	1 ORH04	Alzheimer's diseas
17	91.5	29.1	123	2 A29652	inter-alpha-trypsi
18	91.5	29.1	352	1 T1BOB1	alpha-1-microglobu
19	91	29.0	249	2 T32060	hypothetical prote
20	91	29.0	1965	2 T33216	hypothetical prote
21	89.5	28.5	2167	2 T34395	hypothetical prote
22	88	28.0	57	2 A59204	basic proteinase 1
23	88	28.0	58	1 T1HABK	isoinhibitor K (BP
24	87.5	27.9	61	1 T1VIT1	venom basic protei
25	86.5	27.5	125	1 T1HOB1	alpha-1-microglobu
26	86.5	27.5	349	2 S35708	alpha-1-microglobu
27	86	27.4	2	2 JH0773	Alzheimer's diseas
28	85.5	27.2	337	1 T1PGB1	alpha-1-microglobu
29	85	27.1	2844	2 S28291	hypothetical prote

30	85	27.1	3137	1 A37797	collagen alpha 3(V
31	83.5	26.6	62	2 S19327	venom basic protei
32	82.5	26.3	349	2 S21089	alpha-1-microglobu
33	82	26.1	100	1 T1BO	basic proteinase 1
34	82	26.1	100	1 T1BOSP	spleen basic prote
35	82	26.1	122	1 A55115	uterine plasmin/er
36	82	26.1	265	2 A53390	Kunitz-type protei
37	82	26.1	1391	2 T20406	hypothetical prote
38	81	25.8	60	1 T1BOR	serum basic protei
39	80	25.5	61	1 T1RV2	venom basic protei
40	80	25.5	838	2 T20125	hypothetical prote
41	80	25.5	1208	2 T27822	hypothetical prote
42	80	25.5	1599	2 T16210	hypothetical prote
43	79.5	25.3	64	2 S41399	Kunitz-type protei
44	79	25.2	2225	2 T26063	hypothetical prote
45	78.5	25.0	65	1 T1VIVC	venom basic protei

ALIGNMENTS

RESULT 1
T26859
hypothetical protein Y43F8B.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26859
R:Ainscough, R.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20276
A:Accession: T26859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1743 <WIL>
A:Cross-references: EMBL:AL032623; PIDN:CAA21511.1; CESP:Y43F8B.3
A:Experimental source: clone Y43F8B
C:Genetics:
A:Gene: CESP:Y43F8B.3
A:introns: 65/1; 92/2; 128/1; 229/1; 367/1; 422/1; 486/2; 523/1; 571/1; 628/1; 857/2;

Query Match
Best Local Similarity 36.5%; Score 114.5; DB 2; Length 1743;
Matches 22; Conservative 8; Mismatches 21; Indels 5; Gaps 2;

QY 3 PDVQACTGPF---TSPHLY-LMHYDPRGCGMTFFARGCDGARGFETVACQAC 53
DB 71 PGQSCITLPRQIGTGPRYIPRWYNVPRGRCELFYWGCGGNGNNRQTFOTCOSTC 126

RESULT 2
JG0185
hepatocyte growth factor activator inhibitor type 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0185
R:Itch, H.; Kataoka, H.; Hamasuna, R.; Kitamura, N.; Koono, M.
Biochem. Biophys. Res. Commun. 255, 740-748, 1999
A:Title: Hepatocyte growth factor activator inhibitor type 2 lacking the first kunitz
A:Reference number: JG0185; MUID:99160423
A:Accession: JG0185
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-252 <ITO>
A:Cross-references: GB:AF099016
C:Superfamily: animal Kunitz-type proteinase inhibitor homology
F:133-183/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Query Match
Best Local Similarity 33.1%; Score 104; DB 2; Length 252;
Matches 19; Conservative 9; Mismatches 23; Indels 2; Gaps 1;

A:Accession: A53110
A:Molecule type: protein
A:Residues: 45-57 <CAL1>
R:Veit, H.; Koegler, M.; Gebhard, W.
FEBS Lett. 245, 137-140, 1989
A:Title: The domain structure of the inhibitor subunit of human inter-alpha-trypsin inhibitor
A:Reference number: S03552; MUID:89171290
A:Accession: S03552
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 206-352 <VE12>
R:Malik, N.; Baldyck, M.; Maes, P.; Capon, C.; Mizon, C.; Han, K.K.; Tartar, A.; Fourme
Biol. Chem. Hoppe-Seyler 373, 1009-1018, 1992
A:Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolation
A:Reference number: S28928; MUID:93039735
A:Accession: S28930
A:Status: preliminary
A:Molecule type: protein
A:Residues: 206-215 <MA1>
R:Morelle, W.; Capon, C.; Baldyck, M.; Sautiere, P.; Kounach, M.; Michalski, C.; Fourme
Eur. J. Biochem. 221, 881-888, 1994
A:Title: Chondroitin sulphate covalently cross-links the three polypeptide chains of inter-
A:Reference number: S43466; MUID:94229087
A:Accession: S43466
A:Status: preliminary
A:Molecule type: protein
A:Residues: 206-221 <MO>
R:Wienlewska, H.G.; Bursens, W.H.; Oppenheim, J.D.; Vilcek, J.
Biochemistry 33, 7423-7429, 1994
A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable complex
A:Reference number: A53642; MUID:94271799
A:Accession: A53642
A:Status: preliminary
A:Molecule type: protein
A:Residues: 206-217 <WIS>
R:Calero, M.; Mendez, E.; Garcia, E.
Biochem. Biophys. Acta 1249, 91-99, 1995
A:Title: Expression of the human complex-forming glycoprotein HC (alpha-1-microglobulin)
A:Reference number: S55688; MUID:95284116
A:Accession: S55688
A:Molecule type: protein
A:Residues: 20-24 <CAL2>
R:Bourguignon, J.; Diarra-Mehrour, M.; Sesboue, R.; Fraín, M.; Sala-Trepat, J.M.; Martini
Biochem. Biophys. Res. Commun. 131, 1146-1153, 1985
A:Title: Human inter-alpha-trypsin-inhibitor: characterization and partial nucleotide sequence
A:Reference number: I52208; MUID:86025577
A:Accession: I52208
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 302-352 <BOU>
A:Cross-references: GB:M11562; NID:g186587; PIDN:AAA59194.1; PID:g307077
R:Mojica, E.G.C.; van den Berg, M.; van der Linden, I.K.; Poort, S.R.; Cupers, R.; Bert
Biochem. J. 311, 753-759, 1995
A:Title: Factor IX zutphen: a Cys(18) -> Arg mutation results in formation of a heterodimer
A:Reference number: S59509; MUID:96067589
A:Accession: S59509
A:Molecule type: protein
A:Residues: 27-35, 'Y', '37 <MOJ>
R:Altman, F.; Mizon, J.; Khan, S.R.
Eur. J. Biochem. 236, 984-990, 1996
A:Title: Identification of uronic-acid-rich protein as urinary bikunin, the light chain
A:Reference number: S66434; MUID:96270753
A:Accession: S66434
A:Molecule type: protein
A:Residues: 206-214, 'X', '216-230 <ATM2>
R:Akerstrom, B.; Bratt, T.; Englund, J.J.
FEBS Lett. 362, 50-54, 1995
A:Title: Formation of the alpha(1)-microglobulin chromophore in mammalian and insect cells
A:Reference number: S68728; MUID:95212582
A:Accession: S68728
A:Molecule type: protein
A:Residues: 89-100 <AKE>
R:Jessen, T.E.; Faarvang, K.L.; Ploug, M.

FEBS Lett. 230,195-200, 1988
A:Title: Carbohydrate as covalent crosslink in human inter-alpha-trypsin inhibitor: A
A:Reference number: S02431; MUID:86167187
A:Molecule type: protein
A:Residues: 206-214,'X','216-217 <RES>
R:Lopez, C.; Grubb, A.; Mendez, E.
FEBS Lett. 144, 349-353, 1982
A:Title: Human protein HC displays variability in its carboxyl-terminal amino acid se
A:Reference number: A91304
A:Contents: annotation; variant of alpha-1-microglobulin
A>Note: pooled urine samples contained two forms of this protein, both lacking 57-Lys
R:Hochstrasser, K.; Schonberger, O.L.; Rossmann, I.; Wachter, E.
H:Hoppe-Seyler's Z. physiol. Chem. 362, 1357-1362, 1981
A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte
by affinity chromatography.
A:Reference number: A91698; MUID:82074265
A:Contents: annotation; carbohydrate binding sites
R:Morill, M.; Travis, J.
BIOL. Chem. Hoppe-Seyler 366, 19-21, 1985
A:Title: The reactive site of human inter-alpha-trypsin inhibitor is in the amino-ter
A:Reference number: A90682; MUID:85225940
A:Contents: annotation; inhibitory site
A>Note: in vitro, the first twelve residues of the amino end of the inhibitor appear
C:Comment: Alpha-1-microglobulin and inter-alpha-trypsin inhibitor are proteolyticall
C:Comment: Alpha-1-microglobulin occurs in many physiological fluids including plasma
. It contains at least one brown-yellow chromophore.

Query Match 30.1%; Score 94.5; DB 1; Length 352;
Best Local Similarity 33.3%; Pred. No. 0.00049;
Matches 18; Conservative 8; Mismatches 23; Indels 5; Gaps 1;

OY 5 VOACT-----PTSEPHVLVLMHYDPQRGCMTPFARGCGDGAARGEFTYEACQAC 53
| | | | | :
Db 284 VAACNLPIVRGRCRAFIQLMAFDAYKGCVLFPYGCGCGGNKKFYSEKECHRYC 337

RESULT 5
S04855
Alzheimer's disease amyloid A4 protein - mouse (fragment)
C:Species: Mus musculus domesticus (western European house mouse)
C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 13-Aug-1999
C:Accession: S04855
R:Fukuchi, K.I.; Martin, G.M.; Deeb, S.S.
Nucleic Acids Res. 17, 5396, 1989
A:Title: Sequence of the protease inhibitor domain of the A4 amyloid protein precursor
A:Reference number: S04855; MUID:89345111
A:Accession: S04855
A:Molecule type: mRNA
A:Residues: 1-76 <FUO>
A:Cross-references: EMBL:X15210; NID:q49965; PIDN:CA33280.1; PID:g930133
A>Note: the authors translated the codon GAT for residue 74 as Val
C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinas
C:Keywords: alternative splicing; serine proteinase inhibitor; transmembrane protein
F:3-53/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match 29.9%; Score 94; DB 2; Length 76;
Best Local Similarity 37.5%; Pred. No. 0.00013;
Matches 18; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

OY 6 QACPTSPHLVLMHYDPQRGCMTPFARGCGDGAARGEFTYEACQAC 53
| | | | | :
Db 6 QAETPCRAMISRNYFDYTBGKCVPFFYGGCGGNRNNDTEYCMAVC 53

RESULT 6
A32282
Alzheimer's disease amyloid beta protein precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 17-Aug-1989 #sequence_revision 17-Aug-1989 #text_change 13-Aug-1999
C:Accession: A32282

A:Molecule type: protein
A:Residues: 'A', 24-33, 'X', 35 <BUE>
A:Accession: C34029
A:Molecule type: protein
A:Residues: 47-50, 'X', 52-53 <BU2>
A:Accession: B34029
A:Molecule type: protein
A:Residues: 133, 'X', 135-137, 'X', 139-140, 'X', 142-144, 'X', 146 <BU3>
A:Gene: GDB:TFPI2
A:Cross-References: GDB:354485
C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor
C:Keywords: anticoagulant; blood coagulation; duplication; glycoprotein; heparin binding
F:1-22/Domain: signal sequence #status predicted <Str>
F:23-235/Product: tissue factor pathway inhibitor-2 #status predicted <MAT>
F:36-86/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:96-149/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
F:158-208/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>

RESULT 10
S06678
Alzheimer's disease amyloid beta protein - rhesus macaque (fragment)

C:Species: Macaca mulatta (rhesus macaque)
C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 13-Aug-1999

C:Accession: S06678

R:Koo, E.H.; Sisodia, S.S.; Price, D.L.

submitted to the EMBL Data Library, July 1989

A:Reference number: S06678

A:Accession: S06678

A:Molecule type: mRNA

A:Residues: 1-76 <KOO>

A:Cross-references: EMBL:X15985; NID:g38080; PID:CAA34116.1; PID:g930135

C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinase

C:Keywords: alternative splicing; serine proteinase inhibitor

F:3-53/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match

Best Local Similarity 29.3%; Score 92; DB 2; Length 76;

Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 6 QACTGPTSPHLVLMHYDPORGCMTPPARCGDCAARGFETYEACQAC 53

DB 6 QAGTGPCRAMISRWYFDTEGKCAPFFYGGCGGNRNFTDEYCMAYC 53

RESULT 11

S41082

amyloid precursor protein homolog - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 25-Dec-1994 #sequence_revision 03-Aug-1995 #text_change 29-Aug-1997

C:Accession: S41082

R:Peterson, L.C.; Bjorn, S.E.; Norris, F.; Norris, K.; Sprecher, C.; Foster, D.C.

FEBS Lett. 338, 53-57, 1994

A:Title: Expression, purification and characterization of a kunitz-type protease inhibitor

A:Reference number: S41082; MOID:94139895

A:Accession: S41082

A:Molecule type: DNA

A>Status: preliminary

A:Residues: 1-111 <PEU>

C:Superfamily: animal kunitz-type proteinase inhibitor homology

F:59-109/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match

Best Local Similarity 29.3%; Score 92; DB 2; Length 111;

Matches 16; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

QY 6 QACTGPTSPHLVLMHYDPORGCMTPPARCGDCAARGFETYEACQAC 53

DB 62 EAMTGPCRAVMRWYFDLSKSKCVRFYGGCGGNRNFTSEDIYCMAYC 109

RESULT 12

A32761

hypothetical Alzheimer's disease amyloid beta protein, Alu-containing clone - human (first

C:Species: Homo sapiens (man)

C>Date: 29-Jan-1990 #sequence_revision 10-Apr-1996 #text_change 10-Apr-1996

C:Accession: A32761

R:de Sauvage, F.; Octave, J.N.

Science 245, 651-653, 1989

A:Title: A novel mRNA of the A4 amyloid precursor gene coding for a possibly secreted pr

A:Reference number: A32761; MOID:89346754

A:Accession: A32761

A:Molecule type: mRNA

A:Residues: 1-484 <DES>

A:Cross-references: GB:M28373

A:Note: the authors translated the codon ATG for residue 433 as Ieu

C:Comment: This is the hypothetical translation of a sequence believed to contain clonin

C:Keywords: cloning artifact

Query Match

Best Local Similarity 29.3%; Score 92; DB 4; Length 484;

Matches 18; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 6 QACTGPTSPHLVLMHYDPORGCMTPPARCGDCAARGFETYEACQAC 53

DB 215 QAGTGPCRAMISRWYFDTEGKCAPFFYGGCGGNRNFTDEYCMAYC 262

RESULT 13

A49974

beta-amyloid precursor protein 2 homolog APLP2 - mouse

C:Species: Mus musculus (house mouse)

C>Date: 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999

C:Accession: A49974

R:Slunt, H.H.; Thnikaran, G.; Von Koch, C.; Lo, A.C.; Tanzi, R.E.; Sisodia, S.S.

J. Biol. Chem. 269, 2637-2644, 1994

A:Title: Expression of a ubiquitous, cross-reactive homologue of the mouse beta-amylo

A:Reference number: A49974; MOID:94132029

A:Accession: A49974

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-751 <SID>

A:Cross-references: GB:U15571; NID:g558467; PID:AAA50603.1; PID:g558468

A:Note: sequence extracted from NCBI backbone (NCBI:P144636)

C:Superfamily: Alzheimer's disease amyloid beta protein; animal kunitz-type proteinase

F:310-360/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match

Best Local Similarity 29.3%; Score 92; DB 2; Length 751;

Matches 16; Conservative 8; Mismatches 24; Indels 0; Gaps 0;

QY 6 QACTGPTSPHLVLMHYDPORGCMTPPARCGDCAARGFETYEACQAC 53

DB 313 EAMTGPCRAVMRWYFDLSKSKCVRFYGGCGGNRNFTSEDIYCMAYC 360

RESULT 14

A49321

amyloid beta (A4) homolog 2 precursor - human

N:Alternate names: CDE1-binding protein

C:Species: Homo sapiens (man)

C>Date: 24-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999

C:Accession: A49321; S34644; S40519

R:Sprecher, C.A.; Grant, F.J.; Grimm, G.; O'Hara, P.J.; Norris, K.; Foster

Biochemistry 32, 4481-4486, 1993

A:Title: Molecular cloning of the cDNA for a human amyloid precursor protein homolog:

A:Reference number: A49321; MOID:93250009

A:Accession: A49321

A:Molecule type: preliminary

A>Status: preliminary

A:Residues: 1-763 <SPR>

A:Cross-references: GB:S60099; NID:g300168; PID:AA60589.1; PID:g300169

A:Experimental source: placenta

A:Note: sequence extracted from NCBI backbone (NCBI:N131198, NCBI:P131199)

A:Note: expression was shown in placenta, brain, heart, lung, liver, and kidney

R:yon der Kammer, H.; Klaudiny, J.; Hanes, J.; Schetel, K.H.

submitted to the EMBL Data Library, April 1993

A:Description: The human homologue of the murine CDE1 binding protein is an amyloid p

A:Reference number: S34644

A:Accession: S34644

A:Molecule type: mRNA

A:Residues: 1-763 <SPR>

A:Cross-references: EMBL:Z22572; NID:g394763; PID:CAA80295.1; PID:g394764

R:Masco, W.; Gurubhagavata, S.; Paradis, M.; Romano, D.M.; Sisodia, S.S.; Hyman, B.

Nature Genet. 5, 95-99, 1993

A:Title: Isolation and characterization of APLP2 encoding a homologue of the Alzheimer

A:Reference number: S40519; MOID:94035131

A:Accession: S40519

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-763 <WAS>

A:Cross-references: GB:I27631; NID:g450391; PID:AA41701.1; PID:g450392

C:Genetics:

A:Gene: GDB:APLP2; APLP2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:06 ; Search time 69.26 Seconds
(without alignments)
17,220 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_351

Perfect score: 314
Sequence: 1 CLPDVQACGTSPHVLVLMH.....RGCDGAARGFETVACQDAC 53

Scoring table:

BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents-AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
7: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*
8: /cgn2_6/ptodata/2/1aa/6E.COMB.pep:*
9: /cgn2_6/ptodata/2/1aa/6F.COMB.pep:*
10: /cgn2_6/ptodata/2/1aa/6G.COMB.pep:*
11: /cgn2_6/ptodata/2/1aa/6H.COMB.pep:*
12: /cgn2_6/ptodata/2/1aa/6I.COMB.pep:*
13: /cgn2_6/ptodata/2/1aa/6J.COMB.pep:*
14: /cgn2_6/ptodata/2/1aa/6K.COMB.pep:*
15: /cgn2_6/ptodata/2/1aa/6L.COMB.pep:*
16: /cgn2_6/ptodata/2/1aa/6M.COMB.pep:*
17: /cgn2_6/ptodata/2/1aa/6N.COMB.pep:*
18: /cgn2_6/ptodata/2/1aa/6O.COMB.pep:*
19: /cgn2_6/ptodata/2/1aa/6P.COMB.pep:*
20: /cgn2_6/ptodata/2/1aa/6Q.COMB.pep:*
21: /cgn2_6/ptodata/2/1aa/6R.COMB.pep:*
22: /cgn2_6/ptodata/2/1aa/6S.COMB.pep:*
23: /cgn2_6/ptodata/2/1aa/6T.COMB.pep:*
24: /cgn2_6/ptodata/2/1aa/6U.COMB.pep:*
25: /cgn2_6/ptodata/2/1aa/6V.COMB.pep:*
26: /cgn2_6/ptodata/2/1aa/6W.COMB.pep:*
27: /cgn2_6/ptodata/2/1aa/6X.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
perfect score than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.5	32.3	70	1	US-07-791-213D-5
2	101.5	32.3	70	1	US-08-293-150A-5
3	101.5	32.3	91	1	US-07-791-213D-89
4	101.5	32.3	91	1	US-07-972-387-2
5	101.5	32.3	91	1	US-08-431-412-2
6	101.5	32.3	91	1	US-08-057-971-2
7	101.5	32.3	91	1	US-08-293-150A-89
8	101.5	32.3	68	1	US-07-972-387-37
9	100.5	32.0	68	1	US-07-972-387-40
10	100.5	32.0	68	1	US-08-431-412-37
11	100.5	32.0	68	1	US-08-431-412-40
12	100.5	32.0	68	1	US-08-057-971-37
13	100.5	32.0	68	1	US-08-057-971-40
14	100.5	32.0	89	1	US-07-972-387-10
15	100.5	32.0	89	1	US-08-431-412-10
16	100.5	32.0	89	1	US-08-057-971-10
17	100.5	32.0	101	1	US-07-972-387-22
18	100.5	32.0	101	1	US-07-972-387-26
19	100.5	32.0	101	1	US-08-431-412-22
20	100.5	32.0	101	1	US-08-431-412-28
21	100.5	32.0	101	1	US-08-057-971-22
22	100.5	32.0	101	1	US-08-057-971-28
23	99.5	31.7	89	1	US-07-972-387-8
24	99.5	31.7	89	1	US-08-431-412-8
25	99.5	31.7	89	1	US-08-057-971-8
26	99	31.5	61	2	US-08-829-876-158
27	98.5	31.4	58	3	US-08-676-124-130

28	98.5	31.4	58	3	US-09-414-878-130	Sequence 130, App
29	98.5	31.4	58	3	US-09-240-136-130	Sequence 130, App
30	98.5	31.4	89	1	US-07-972-387-16	Sequence 16, App1
31	98.5	31.4	89	1	US-08-431-412-16	Sequence 16, App1
32	98.5	31.4	89	1	US-08-057-971-16	Sequence 16, App1
33	98.5	31.4	61	2	US-08-829-876-141	Sequence 141, App1
34	98.5	31.2	61	2	US-08-829-876-141	Sequence 141, App1
35	97	30.9	57	2	US-08-829-876-141	Sequence 141, App1
36	97	30.9	61	2	US-08-829-876-141	Sequence 141, App1
37	97	30.9	61	2	US-08-829-876-141	Sequence 141, App1
38	96.5	30.7	68	1	US-07-972-387-35	Sequence 35, App1
39	96.5	30.7	68	1	US-07-972-387-42	Sequence 42, App1
40	96.5	30.7	68	1	US-08-431-412-35	Sequence 35, App1
41	96.5	30.7	68	1	US-08-431-412-42	Sequence 42, App1
42	96.5	30.7	68	1	US-08-057-971-35	Sequence 35, App1
43	96.5	30.7	68	1	US-08-057-971-42	Sequence 42, App1
44	96.5	30.7	68	2	US-08-235-515A-22	Sequence 22, App1
45	96.5	30.7	68	2	US-08-235-515A-23	Sequence 23, App1

ALIGNMENTS

RESULT 1
US-07-791-213D-5
Sequence 5, Application US/07791213D
Patent No. 5409895
GENERAL INFORMATION:
APPLICANT: MORISHITA, Hideaki
APPLICANT: KANAMORI, Toshinori
APPLICANT: NOBUHARA, Masahiro
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
NUMBER OF SEQUENCES: 108
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07791, 213D
FILING DATE: 13-NOV-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-306745
FILING DATE: 13-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Neuth, Donna M
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 029650-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 70 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-791-213D-5
Query Match 32.3%; Score 101.5; DB 1; Length 70;
Best Local Similarity 35.2%; Pred. No. 1.8e-05;

RESULT 3
US-07-791-213D-89
; Sequence 89, Application US/07791213D
; Patent No. 5409895
; GENERAL INFORMATION:
; APPLICANT: MORISHITA, Hideaki

```
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: PatentIn Release #1.0, Version #1.25
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/07/972,387
5 FILING DATE: 19921105
6 CLASSIFICATION: 435
7 ATTORNEY/AGENT INFORMATION:
8 NAME: Murphy Jr., Gerald M.
9 REGISTRATION NUMBER: 28,977
10 REFERENCE/DOCKET NUMBER: 1110-124P
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: 703-241-1300
13 TELEFAX: 703-241-2848
14 TELEX: 248345
15 INFORMATION FOR SEQ ID NO: 2:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 91 amino acids
18 TYPE: AMINO ACID
19 TOPOLOGY: linear
20 MOLECULE TYPE: protein
21
22 JS-07-972-387-2

```

Query Match	32.3%	Score 101.5	DB 1	Length 91
Best Local Similarity	35.2%	Pred. No. 2.4e-05		
Matches	19	Conservative	8	Mismatches 22; Indels 5; Gaps 1;
Oy	5	VOACT-----GPTSPHYLMHTDPORGCGMTTPAGCCDGAANGETFEYECQAC	53	
DB	23	VAACNLPVYRGCGRAFTQTMADAVKAGKRVLEPYGCGCGNGKMGFSEKRCRYC	76	

RESULT 5
 US-08-431-412-2
 : Sequence 2, Application us/08431412
 : Patent No. 5589360
 : GENERAL INFORMATION:
 : APPLICANT: Morishita, Hideaki
 : APPLICANT: Kanamori, Toshinori
 : APPLICANT: No. 5589360uhara, Masahiro
 : TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the
 : TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for
 : TITLE OF INVENTION: Producing the Same
 : NUMBER OF SEQUENCES: 76
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Birch, Stewart, Kolasch & Birch
 : STREET: 301 N. Washington St.
 : CITY: Falls Church
 : STATE: Virginia
 : COUNTRY: USA
 : ZIP: 22046-0747
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/431,412
 : FILING DATE: 28-APR-1995
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/972,387
 : FILING DATE: 05-NOV-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Murphy Jr., Gerald M.
 : REGISTRATION NUMBER: 28,977
 : REFERENCE/DOCKET NUMBER: 1110-124P
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 703-241-1300
 : TELEFAX: 703-241-2848
 : TELEX: 248345
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:

```

;      LENGTH:  91 amino acids
;      TYPE:    amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE:  protein
US-08-431-412-2

```

Query Match	32.3%	Score	101.5	DB 1	Length	91			
Best Local Similarity	35.2%	Pred. No.	2.4e-05						
Matches	19	Conservative	8	Mismatches	22	Indels	5	Gaps	1

QY 5 VACT-----GPTSPHLVLMHYDPQRGCMTFPARGCDGAARGFETYEACQDAC 53
| | : : : : : : : : : :
Db 23 VACNLPIVRGPCRAFTLMAFDVAVKGVLFPPYGCGGNGNKFESEKECREYC 76

```

RESULT 6
US-08-057-971-2
; Sequence 2, Application US/08057971
; Patent No. 5679770
; GENERAL INFORMATION:
; APPLICANT: Morishita, Hideaki
; APPLICANT: Kanamori, Toshinori
; APPLICANT: No. 5679770uhara, Masahiro
; TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the
; TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for
; TITLE OF INVENTION: Producing the Same
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/057,971
; FILING DATE: 06-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1110-129P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-057-971-2

Query Match 32.3%; Score 101.5; DB 1; Length 91;
Best Local Similarity 35.2%; Pred. No. 2,4e-05;
Matches 19; Conservative 8; Mismatches 22; Indels 5; Gaps 1;

QY 5 VQACT-----GPTSPHLVLMHYDPQRGCGMTTPARCGDGAARFEFYEAQCAQC 53
DB 23 VAAACNLPVIRGCGRAFIQLMFDVAVKGKCVLPFGGCGGNGMNFESKECREYCV 76

RESULT 7
US-08-293-150A-89
; Sequence 89, Application US/08293150A

```

Patent No. 5792629
GENERAL INFORMATION:
APPLICANT: MORISHITA, Hideaki
APPLICANT: KANAMORI, Toshiro
APPLICANT: NOBUHARA, Masahiro
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
TITLE OF INVENTION: TREATING USING THE SAME
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DONALD, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/293,150A
FILING DATE: 13-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/791,213
FILING DATE: 13-NOV-1990
PRIOR APPLICATION DATA: JP 2-306745
APPLICATION NUMBER: JP 2-306745
FILING DATE: 13-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 029650-049
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 91 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-293-150A-89

Query Match 32.3%; Score 101.5; DB 1; Length 91;
Best Local Similarity 35.2%; Pred. No. 2.4e-05;
Matches 19; Conservative 8; Mismatches 22; Indels 5; Gaps 1;

QY 5 VQACT-----GPTSPHLVLMHYDPQGGCTPPARGCGAARGPFTYEACQAC 53
DB 23 VAAACNLPYAGPCRAFIQIMAFDAVKCKVLPFYGGCGGNGNKFSEKRECYC 76

RESULT 8
US-07-972-387-37
Sequence 37, Application US/07972387
Patent No. 5451659
GENERAL INFORMATION:
APPLICANT: MORISHITA, Hideaki
APPLICANT: KANAMORI, Toshiro
APPLICANT: NOBUHARA, Masahiro
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
TITLE OF INVENTION: TREATING USING THE SAME
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church

STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/972,387
FILING DATE: 19921105
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1110-124P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 68 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-07-972-387-37

Query Match 32.0%; Score 100.5; DB 1; Length 68;
Best Local Similarity 36.5%; Pred. No. 2.3e-05;
Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

QY 2 LPVQACTGPTSPHLVLMHYDPQGGCTPPARGCGAARGPFTYEACQAC 53
DB 5 LPVQ---GPCRAFIKIMADAVKCKVLPFYGGCGGNGNKFSEKRECYC 53

RESULT 9
US-07-972-387-40
Sequence 40, Application US/07972387
Patent No. 5451659
GENERAL INFORMATION:
APPLICANT: MORISHITA, Hideaki
APPLICANT: KANAMORI, Toshiro
APPLICANT: NOBUHARA, Masahiro
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
TITLE OF INVENTION: TREATING USING THE SAME
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/972,387
FILING DATE: 19921105
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977

```
REFERENCE/DOCKET NUMBER: 1110-124P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 68 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Eschericia coli
US-07-972-387-40

Query Match          32.0%; Score 100.5; DB 1; Length 68;
Best Local Similarity 36.5%; Pred. No. 2.3e-05;
Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

Oy 2 LPDVACTGPTSPHLVLMHYDPQGGCMTPPARGCDGAARGFTYBACQAC 53
Db 5 LPVQ---GPCRAFIKLMFADAVKGCVLFPYGGCGNGNKFSEKRECREYC 53

RESULT 10
US-08-431-412-37
Sequence 37, Application US/08431412
Patent No. 5589360
GENERAL INFORMATION:
APPLICANT: Morishita, Hideaki
APPLICANT: Kanamori, Toshinori
APPLICANT: No. 5589360uhara, Masahiro
TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the
TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for
TITLE OF INVENTION: Producing the Same
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,412
FILING DATE: 28-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/972,387
FILING DATE: 05-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1110-124P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 68 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
```

```
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Eschericia coli
US-08-431-412-37

Query Match          32.0%; Score 100.5; DB 1; Length 68;
Best Local Similarity 36.5%; Pred. No. 2.3e-05;
Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

Oy 2 LPDVACTGPTSPHLVLMHYDPQGGCMTPPARGCDGAARGFTYBACQAC 53
Db 5 LPVQ---GPCRAFIKLMFADAVKGCVLFPYGGCGNGNKFSEKRECREYC 53

RESULT 11
US-08-431-412-40
Sequence 40, Application US/08431412
Patent No. 5589360
GENERAL INFORMATION:
APPLICANT: Morishita, Hideaki
APPLICANT: Kanamori, Toshinori
APPLICANT: No. 5589360uhara, Masahiro
TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the
TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for
TITLE OF INVENTION: Producing the Same
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,412
FILING DATE: 28-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/972,387
FILING DATE: 05-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1110-124P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 68 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Eschericia coli
US-08-431-412-40

Query Match          32.0%; Score 100.5; DB 1; Length 68;
Best Local Similarity 36.5%; Pred. No. 2.3e-05;
Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

Oy 2 LPDVACTGPTSPHLVLMHYDPQGGCMTPPARGCDGAARGFTYBACQAC 53
Db 5 LPVQ---GPCRAFIKLMFADAVKGCVLFPYGGCGNGNKFSEKRECREYC 53
```

DB 5 LPIVE---GPCRAFIKIMAFDAVKGKCVLPYGGCGNGNKFSEKREXC 53

RESULT 12

US-08-057-971-37
Sequence 37, Application US/08057971
Patent No. 5679770

GENERAL INFORMATION:

APPLICANT: Morishita, Hideaki

APPLICANT: Kanamori, Toshiro

APPLICANT: No. 5679770uhara, Masahiro

TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the

TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for

NUMBER OF SEQUENCES: 81

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/057,971

FILING DATE: 06-MAY-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1110-129P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

TELEX:

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 68 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: C-terminal

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

US-08-057-971-37

Query Match 32.0%; Score 100.5; DB 1; Length 68;

Best Local Similarity 36.5%; Pred. No. 2.3e-05;

Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

DB 5 LPIVE---GPCRAFIKIMAFDAVKGKCVLPYGGCGNGNKFSEKREXC 53

RESULT 13

US-08-057-971-40

Sequence 40, Application US/08057971

Patent No. 5679770

GENERAL INFORMATION:

APPLICANT: Morishita, Hideaki

APPLICANT: Kanamori, Toshiro

APPLICANT: No. 5679770uhara, Masahiro

TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the

TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for

NUMBER OF SEQUENCES: 81

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/057,971

FILING DATE: 06-MAY-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1110-129P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

TELEX:

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 68 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: C-terminal

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

US-08-057-971-40

Query Match 32.0%; Score 100.5; DB 1; Length 68;

Best Local Similarity 36.5%; Pred. No. 2.3e-05;

Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

DB 5 LPIVE---GPCRAFIKIMAFDAVKGKCVLPYGGCGNGNKFSEKREXC 53

RESULT 14

US-07-972-387-10

Sequence 10, Application US/07972387

Patent No. 5451659

GENERAL INFORMATION:

APPLICANT: Morishita, Hideaki

APPLICANT: Kanamori, Toshiro

APPLICANT: No. 5451659uhara, Masahiro

TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the

TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for

NUMBER OF SEQUENCES: 76

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: 301 N. Washington St.

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22046-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/972,387

FILING DATE: 19921105

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1110-124P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 89 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-972-387-10

Query Match 32.0%; Score 100.5; DB 1; Length 89;
Best Local Similarity 36.5%; Pred. No. 3,1e-05;
Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

QY 2 LPDVQACTGPTSPHLVLMHYDPORGCGMTFPAKCDGAARGFETYACQAC 53
DB 26 LPIVE---GPCRAFIQLMAFDVAKGKCVLPYGGCGGNGNKFESEREKREYC 74

RESULT 15
US-08-431-412-10

Sequence 10, Application US/08431412
Patent No. 5589360

GENERAL INFORMATION:

APPLICANT: Morishita, Hideaki
APPLICANT: Kanamori, Toshinori
APPLICANT: No. 5589360uhara, Masahiro
TITLE OF INVENTION: Polypeptide, DNA Fragment Encoding the
TITLE OF INVENTION: Same, Drug Composition Containing the Same and Process for
TITLE OF INVENTION: Producing the Same
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSER: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,412

FILING DATE: 28-APR-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/972,387

FILING DATE: 05-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1110-124P

TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300

TELEFAX: 703-241-2848

TELEX: 248345

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 89 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-431-412-10

Query Match 32.0%; Score 100.5; DB 1; Length 89;
Best Local Similarity 36.5%; Pred. No. 3,1e-05;
Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

QY 2 LPDVQACTGPTSPHLVLMHYDPORGCGMTFPAKCDGAARGFETYACQAC 53
DB 26 LPIVE---GPCRAFIQLMAFDVAKGKCVLPYGGCGGNGNKFESEREKREYC 74

Search completed: February 26, 2002, 01:28:06
Job time: 360 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:43 ; Search time 162.6 Seconds
(without alignments)
24.144 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_351
Perfect score: 314
Sequence: 1 CLPDVQACGCPPTSPHLVLMH.....RGCDGAARCFETYEACQAC 53

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneeqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneeqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneeqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneeqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneeqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneeqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneeqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneeqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneeqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneeqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneeqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneeqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneeqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneeqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneeqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneeqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneeqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneeqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneeqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneeqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneeqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneeqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102	32.5	61	21	AAV68111
2	101.5	32.3	91	13	AAAR23942
3	101.5	32.3	91	14	AAAR36751
4	100.5	32.0	101	14	AAAR36761
5	99.5	31.7	89	14	AAAR36756
6	99	31.5	58	17	AAAR9176
7	99	31.5	61	21	AAV68109
8	98.5	31.4	58	16	AAAR78563
9	98.5	31.4	101	14	AAAR36764
10	98	31.2	61	17	AAAW18397
11	98	31.2	61	17	AAAW18401

12	98	31.2	61	21	AAV68092	Kunitz protease in
13	98	31.2	61	21	AAV68115	Kunitz protease in
14	97.5	31.1	58	17	AAAR9205	Genetically engine
15	97.5	31.1	89	14	AAAR36754	Factor x inhibitor
16	97	30.9	57	17	AAAW18395	KPI(1-57) EID, M15
17	97	30.9	57	17	AAAW18410	KPI(1 to 57) S171
18	97	30.9	57	21	AAV68079	Kunitz protease in
19	97	30.9	61	17	AAV68039	KPI(-4 to 57) R40A
20	97	30.9	61	17	AAV68402	KPI(-4 to 57) M15L
21	97	30.9	61	17	AAV68416	KPI(-4 to 57) S171
22	97	30.9	61	21	AAV68117	Kunitz protease in
23	97	30.9	61	21	AAV68145	Kunitz protease in
24	97	30.9	123	12	AAAR4760	Amyloid precursor
25	96.5	30.7	51	16	AAAR68873	Protease (esp. ela
26	96.5	30.7	66	15	AAAR61160	Elastase inhibitor
27	96.5	30.7	66	15	AAAR61161	Elastase inhibitor
28	96.5	30.7	66	15	AAAR61163	Elastase inhibitor
29	96.5	30.7	66	15	AAAR61164	Elastase inhibitor
30	96.5	30.7	68	15	AAAR65479	The amino acid seq
31	96.5	30.7	68	16	AAAR68874	Protease (esp. ela
32	96.5	30.7	72	13	AAAR23940	Factor x inhibitor
33	96.5	30.7	101	14	AAAR36762	Factor x inhibitor
34	96.5	30.7	101	14	AAAR36766	Factor x inhibitor
35	96.5	30.7	101	15	AAAR65484	Fusion protein con
36	96.5	30.7	101	15	AAAR65485	Fusion protein con
37	96.5	30.7	101	16	AAAR68879	Protease (esp. ela
38	96.5	30.7	147	19	AAAR69522	Genetically engine
39	96	30.6	58	17	AAAR91172	Genetically engine
40	96	30.6	58	17	AAAR91174	Genetically engine
41	95.5	30.4	70	11	AAAR08292	Human btkunlin doma
42	95.5	30.4	91	13	AAAR23943	Plasmin PM576 regl
43	95.5	30.4	101	14	AAAR36760	Factor x inhibitor
44	95	30.3	61	17	AAAW18422	KPI(-4 to 57) S17L
45	95	30.3	61	21	AAV68149	Kunitz protease in

ALIGNMENTS

RESULT 1	AAV68111	standard; Protein; 61 AA.
ID	AAV68111;	
AC	AAV68111;	
XX		
DT		
XX		
DE		
XX		
KW	Kunitz protease inhibitor; KPI, serine protease; kallikrein; plasmin;	
KW	amyloid precursor protein; coagulation factor; blood loss; cardiant;	
KW	cardiopulmonary bypass surgery; anticoagulant; anti-inflammatory;	
KW	anti-arthritic; thrombolytic; antirheumatic; antipneumatic;	
KW	immunosuppressant; pancreatitis; deep vein thrombosis; psoriasis;	
KW	rheumatoid arthritis; myocardial infarction; transplant rejection.	
OS	Homo sapiens.	
OS	Synthetic.	
PN	WO9963090-A2.	
XX		
PD	09-DEC-1999.	
XX		
PF	03-JUN-1999;	99WO-US12276.
XX		
PR	03-JUN-1998;	98US-0087885.
XX		
PA	(SCIO-) SCIOS INC.	
XX		
PI	White RT, Damm D, Lesikar DD, McFadden K, Garrick BL, Lucas AB;	
PI	Pollitt NS, Lam AO;	
XX		
DR	WPI, 2000-105699/09.	

xx	Novel enzyme inhibitors especially used to reduce postoperative
PT	bleeding -
PT	
xx	
PS	Example 4; Fig 64; 151pp; English.
xx	
CC	The present invention describes protease inhibitors that
CC	are analogues of the Kunitz protease inhibitor (KPI) domain of the
CC	amyloid precursor protein. The protease inhibitors can be used used to
CC	treat or prevent disorders associated with increased activity of serine
CC	proteases, specifically blood loss during surgery (particularly
CC	cardiopulmonary bypass surgery where plasma proteases are activated by
CC	contact with surfaces in the heart-lung machine), but also other
CC	conditions such as pancreatitis; deep vein thrombosis; Rheumatoid
CC	arthritis; psoriasis; myocardial infarction; and transplant rejection.
CC	They are also for organ preservation and to promote wound healing. In
CC	vitro the protease inhibitors may be used to inhibit serine proteases
CC	during preparation of cell extracts. The protease inhibitors are based
CC	on a human peptide sequence so are unlikely to be immunogenic, can be
CC	produced at high levels in recombinant expression systems, and can
CC	inhibit a wide range of serine proteases. They are more potent or
CC	specific than known inhibitors. The present sequence represents a KPI
xx	variant which is given in an example from the present invention.
xx	
Sequence	61 Aa:
xx	

Query Match	32.5%	Score 102;	DB 21;	Length 61;
Best Local Similarity	39.6%;	Pred. No. 9.8e-06;		
Matches 19; Conservative	4;	Mismatches 25;	Indels 0;	Gaps 0;

0Y 6 QACTGPTSPHLVLWHYPDQGGCMTFPARCGDGAARGFETYEACQAC 53
 || ||| :- :-:- | | | | | | | | |
Db 10 qaetgpcrahisrwyfdvteqkcapflygcghnrmnfdeeycmavc 57

PT heart disease, thrombosis, arthritis, allergy, shock, etc.
XX
PS Disclosure; Fig 21. 106pp: English.
XX
CC The sequence given is encoded by a portion of the plasmid pM5753
CC between the HindIII and BamHI recognition sites and is the
CC polypeptide of the invention (polypeptide v46E).
CC The DNA sequence encoding v46E was derived by PCR using the primer
CC sequences given in AA024142 and AA024144.
CC Polypeptide v46E is a modified version of a novel polypeptide which
CC comprises the amino acid sequence that constitutes a portion of
CC urinary trypsin inhibitor (UTI). This polypeptide has strong
CC inhibitory activity against proteases such as trypsin, elastase,
CC plasmin kallikrein and Fxa, and can be used in compns. to treat
CC diseases caused by these enzymes, eg. operative stress, multiple organ
CC failure, shock, pancreatitis, ischaemic heart disease, nephritis,
CC hepatic cirrhosis, thrombosis after revascularisation, oedema caused
CC by increased vascular permeability, adult respiratory distress
CC syndrome, rheumatoid arthritis, arthritis and/or allergy, disseminated
CC intravascular coagulation syndrome. It may also be used to prevent
CC general blood coagulation. The peptide shows no antigenicity against
CC humans.
XX
XX Sequence 91 AA;
SQ

Query Match	32.38;	Score 101.5;	DB 13;	Length 91;
Best Local Similarity	35.28;	Pred. No. 1.7e-05;		
Matches 19;	Conservative 8;	Mismatches 22;	Indels 5;	Gaps 1

DQ 5 VQACT-----GPTSPHLVLMHYDPQRGSGMTFPARGCDGAARGFETIYEAQQAC 53
||| ||| : ||| : ||| ||| ||| : ||| :
Db 23 vacnlpivrgpcrafiqlwafdavkscvllfpygsqgnynkfesekcreyc 76

RESULT	2
AAR23942	
ID	AAR23942 standard; Protein; 91 AA

DT 15-NOV-1992 (first entry)
VY

Plasmid pM575B region.

KW Polymerase chain reaction; PCR; UTI; Y46E; HindIII; BamHI.
 YY

Synthetic.

FH	Key	Location/Qualifiers
EM	port+1:30	1 31

Protein	/label= phoA_signal_peptide
FT	32.91

/label= Polypeptide_Y46E

PN EP486001-A.

PD 20-MAY-1992.

PF 13-NOV-1991; 91EP-0119378.

PR 13-NOV-1990; 90JP-0306745.

(MOCH) MOCHIDA PHARM CO LTD,
PA
XX

XX
P1
XXXXXXXXXX, Morishita H, Nobunara M,

DR N-PSDB; AA024152.

PT New polypeptides comprise amino acid sequence of urinary trypsin inhibitor - are protease inhibitors for treating e.g. ischaemic

heart disease, thrombosis, arthritis, allergy, shock, etc.
Disclosure; Fig 21, 106pp: English.

The sequence given is encoded by a portion of the plasmid pm575B between the HindIII and BamHI recognition sites and is the polypeptide of the invention (polypeptide Y46E). The DNA sequence encoding Y46E was derived by PCR using the entire DNA sequence encoding Y46E as template.

Polypeptide Y46E is a modified version of a novel polypeptide which comprises the amino acid sequence that constitutes a portion of urinary trypsin inhibitor (UTI). This polypeptide has strong

plasma kallikrein and FXa, and can be used in compounds to treat diseases caused by these enzymes, eg. operative stress, multiple organ failure, shock, pancreatitis, ischemic heart disease, nephritis, and myocardial infarction.

by increased vascular permeability, adult respiratory distress syndrome, rheumatoid arthritis, arthritis and/or allergy, disseminated intravascular coagulation syndrome. It may also be used to prevent

humans.
Sequence 91 AA;

RESULT	3
AAR36751	
ID	AAR36751 standard; Protein; 91 AA.

AC AAR36751,
VY

DT 16-SEP-1993 (first entry)

Factor X inhibiting peptide, Y42E, coding sequence

KW Factor Xa; urinary trypsin inhibitor; Bikunin; elastase;
KW substitution; mutation; growth; tumor; metastasis;
KW

KW pancreatic disease; rheumatoid arthritis.

OS Escherichia coli, strain JE5505 (PM575B)

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```
ET      Dec 01 1996 14:23:00
ET      /label= Y42E
ET      36 76
```

FT /note= "Claim 1, may opt. be sequentially extended at

line N- and C-terminal by one amino acid to include bases upto position 22 and upto position 91 respectively, may also be extended at the N-terminal by the sequence given in AAR39756*

PN EP543240-A.
VY

PD 26-MAY-1993

PF 06-NOV-1992; 92EP-0119083.

PR	08-NOV-1991;	91JP-0293472.
PR	13-MAY-1993;	92TB-0118386

XX
XX
PA (MOCH) MOCHTDA PHARM CO LTD

PI Kanamori T, Morishita H, Nobuhara M;
 XX WPI: 1993-168945/21.
 DR N-PSDB; AA041869.
 XX
 XX New polypeptide inhibiting protease(s), esp. FXa - used for
 PT treating multiple organ failure, shock, pancreatitis,
 PT disseminated intravascular coagulation, etc.
 XX
 XX Disclosure; Page 45; 130pp; English.
 XX
 CC The sequences given in AAR36751-67 represent polypeptides which have
 CC high factor Xa inhibition activity. These peptides are based on a wild
 CC type sequence which coincides with part of the amino acid sequence of
 CC urinary trypsin inhibitor (UTI) or Btkunins (HI-30). It is different
 CC to both of these proteins however, in its factor Xa inhibiting
 CC activity. Substitutions/mutations of the wild type sequence may
 CC increase factor Xa inhibiting activity, improve secretion of the
 CC polypeptide from the host cell or increase the ability of the protein
 CC to inhibit other proteins, eg. elastase. These properties may also be
 CC effected by supplementing one or more amino acids at the C- and/or
 CC N-terminal of these proteins. These peptides may be used in drug
 CC compositions for the prevention and/or treatment of infestation,
 CC shock, pancreatitis, ischaemic heart disease and rheumatoid arthritis.
 XX
 SQ Sequence 91 AA:
 Query Match 32.3%; Score 101.5; DB 14; Length 91;
 Best Local Similarity 35.2%; Pred. No. 1.7e-05;
 Matches 19; Conservative 8; Mismatches 22; Indels 5; Gaps 1;
 QY 5 VQACT-----GPTSHLVLMHYDPORGCMTPFARCGDGAARGFETYACQAC 53
 Db 23 vaacnlplivgpcratflqlwaIdavkKgcVlfpYgscqgnknkIsekreyc 76
 RESULT 4
 AAR36761
 ID AAR36761 standard; Protein; 101 AA.
 XX
 AC AAR36761;
 XX
 DT 16-SEP-1993 (first entry)
 XX
 DE Factor X inhibiting peptide, R7Q/Q15R/Y42D, coding sequence.
 XX
 KW Factor Xa; inhibition; urinary trypsin inhibitor; Btkunin; elastase;
 KW substitution; mutation; secretion; drug; UTI; infestation; shock;
 KW pancreatitis; ischaemic heart disease; rheumatoid arthritis.
 XX
 OS Escherichia coli, strain JES505 (pm741).
 XX
 XX Key Location/Qualifiers
 FH MISC-difference 42
 FT /label- R7Q
 FT MISC-difference 50
 FT /label- Q15R
 FT MISC-difference 77
 FT /label- Y42D
 FT Protein 24..74
 FT /note- "may opt. be sequentially extended at
 FT the N- and C-terminal by one amino acid to
 FT include bases upto position 34 and upto
 FT position 101 respectively, may also be
 FT extended at the N-terminal by the sequence
 FT given in AAR39758"
 XX
 PN EP543240-A.
 XX
 XX 26-MAY-1993.
 PD
 XX
 PF 06-NOV-1992; 92EP-0119083.

XX
 PR 08-NOV-1991; 91JP-0293472.
 PR 12-MAY-1992; 92JP-0119289.
 XX
 XX (MOCH) MOCHIDA PHARM CO LTD.
 XX
 PI Kanamori T, Morishita H, Nobuhara M;
 XX WPI: 1993-168945/21.
 DR N-PSDB; AA041879.
 XX
 XX New polypeptide inhibiting protease(s), esp. FXa - used for
 PT treating multiple organ failure, shock, pancreatitis,
 PT disseminated intravascular coagulation, etc.
 XX
 XX Disclosure; Page 65; 130pp; English.
 XX
 CC The sequences given in AAR36751-67 represent polypeptides which have
 CC high factor Xa inhibition activity. These peptides are based on a wild
 CC type sequence which coincides with part of the amino acid sequence of
 CC urinary trypsin inhibitor (UTI) or Btkunins (HI-30). It is different
 CC to both of these proteins however, in its factor Xa inhibiting
 CC activity. Substitutions/mutations of the wild type sequence may
 CC increase factor Xa inhibiting activity, improve secretion of the
 CC polypeptide from the host cell or increase the ability of the protein
 CC to inhibit other proteins, eg. elastase. These properties may also be
 CC effected by supplementing one or more amino acids at the C- and/or
 CC N-terminal of these proteins. These peptides may be used in drug
 CC compositions for the prevention and/or treatment of infestation,
 CC shock, pancreatitis, ischaemic heart disease and rheumatoid arthritis.
 XX
 SQ Sequence 101 AA:
 Query Match 32.0%; Score 100.5; DB 14; Length 101;
 Best Local Similarity 36.5%; Pred. No. 2.6e-05;
 Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;
 QY 2 LPDVACTGPTSPHLVLMHYDPORGCMTPFARCGDGAARGFETYACQAC 53
 Db 38 lplvg---gpcratflklwaIdavkKgcVlfpYgscqgnknkIsekreyc 86
 RESULT 5
 AAR36756
 ID AAR36756 standard; Protein; 89 AA.
 XX
 AC AAR36756;
 XX
 DT 16-SEP-1993 (first entry)
 XX
 DE Factor X inhibiting peptide, Y42E, coding sequence.
 XX
 KW Factor Xa; inhibition; urinary trypsin inhibitor; Btkunin; elastase;
 KW substitution; mutation; secretion; drug; UTI; infestation; shock;
 KW pancreatitis; ischaemic heart disease; rheumatoid arthritis.
 XX
 OS Escherichia coli, strain JES505 (pm575C).
 XX
 XX Key Location/Qualifiers
 FH MISC-difference 65
 FT /label- Y42E
 FT Protein 24..74
 FT /note- "may opt. be sequentially extended at
 FT the N- and C-terminal by one amino acid to
 FT include bases upto position 22 and upto
 FT position 89 respectively, may also be
 FT extended at the N-terminal by the sequence
 FT given in AAR39758"
 XX
 PN EP543240-A.
 XX
 XX 26-MAY-1993.
 PD

XX 06-NOV-1992; 92EP-0119083.
 PF 08-NOV-1991; 91JP-0293472.
 XX 12-MAY-1992; 92JP-0119289.
 PR (MOCH) MOCHIDA PHARM CO LTD.
 XX
 XX Kanamori T, Morishita H, Nobuhara M;
 PI
 DR WPI; 1993-168945/21.
 DR N-PSDB; AA041874.
 XX
 PT New polypeptide inhibiting protease(s), esp. FXa - used for
 PT treating multiple organ failure, shock, pancreatitis,
 PT disseminated intravascular coagulation, etc.
 PS
 XX Disclosure; Page 55; 130pp; English.
 XX
 CC The sequences given in AAR36751-67 represent polypeptides which have
 CC high factor Xa inhibition activity. These peptides are based on a wild
 CC type sequence which coincides with part of the amino acid sequence of
 CC urinary trypsin inhibitor (UTI) or Bixunins (HI-30). It is different
 CC to both of these proteins however, in its factor Xa inhibiting
 CC activity. Substitutions/mutations of the wild type sequence may
 CC increase factor Xa inhibiting activity, improve secretion of the
 CC polypeptide from the host cell or increase the ability of the protein
 CC to inhibit other proteins, eg. elastase. These properties may also be
 CC effected by supplementing one or more amino acids at the C- and/or
 CC N-terminal of these proteins. These peptides may be used in drug
 CC compositions for the prevention and/or treatment of intestinal,
 CC shock, pancreatitis, ischaemic heart disease and rheumatoid arthritis.
 CC
 XX Sequence 89 AA;

Query Match 31.7%; Score 99.5; DB 14; Length 89;
 Best Local Similarity 36.5%; Pred. No. 3e-05;
 Matches 19; Conservative 9; Mismatches 21; Indels 3; Gaps 1;

QY 2 LPPVOACTGTSPLVIMHYDPQRCGCMTPPARGCDGAARGFETYEACQAC 53
 DB 26 Lpivr---gpcrafiqlwaidavkxcvllfpygscgqgnkfkfesekecreyc 74

RESULT 6
 ID AAR99176 standard; protein; 58 AA.
 AC AAR99176;

DT 12-FEB-1997 (first entry)

DE Genetically engineered apocrotin-like Kunitz domain (DPI.2.3).

XX Aprotinin; Kunitz domain; human neutrophil elastase; hNE;
 KW connective tissue; alpha 1 protease inhibitor; API; neutrophil;
 KW alpha 1 antitrypsin; respiratory disorder; cystic fibrosis;
 KW smokers emphysema.

XX Synthetic.

XX WO9620278-A2.

PD 04-JUL-1996.

PF 15-DEC-1995; 95WO-US16349.

PR 16-DEC-1994; 94US-0358160.

XX (PROF-) PROTEIN ENG CORP.

PA Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W;
 PI

PI Roberts BL;
 XX WPI; 1996-321851/32.
 XX
 PT New engineered inhibitors of human neutrophil elastase - contg.
 PT aprotinin-like Kunitz domain for treating, e.g. cystic fibrosis or
 PT other respiratory disorders
 PS
 XX Claim 1; Page 50; 105pp; English.
 CC Genetically engineered human derived Kunitz domains can be used to
 CC inhibit human neutrophil elastase, an enzyme involved in the
 CC elimination of pathogens and the restructuring of connective tissue.
 CC In cases of reduction of the circulating alpha-1-protease inhibitor
 CC (API or alpha 1 antitrypsin), or the inactivation of API by oxidation
 CC (smokers emphysema), extensive destruction of the lung tissue may
 CC result from uncontrolled elastolytic activity of human neutrophil
 CC elastase. Other respiratory disorders such as cystic fibrosis are
 CC thought to be caused by human neutrophil elastase release by
 CC neutrophils. The genetically engineered human derived Kunitz
 CC domains can be used to treat such respiratory disorders. See
 CC AAR99146-R99211.
 XX
 XX Sequence 58 AA;

Query Match 31.5%; Score 99; DB 17; Length 58;
 Best Local Similarity 42.2%; Pred. No. 2.2e-05;
 Matches 19; Conservative 4; Mismatches 22; Indels 0; Gaps 0;

QY 9 TGPTSPHLVIMHYDPQRCGCMTPPARGCDGAARGFETYEACQAC 53
 DB 11 tgciafflryyydrytgcscqtfyggcmgnannfyweacddac 55

RESULT 7
 ID AAY68109 standard; protein; 61 AA.
 AC AAY68109;

DT 13-APR-2000 (first entry)

DE Kunitz protease inhibitor variant DD101.

XX Kunitz protease inhibitor; KPI; serine protease; kallikrein; plasmin;
 KW amyloid precursor protein; coagulation factor; blood loss; cardiant;
 KW cardiopulmonary bypass surgery; anticoagulant; anti-inflammatory;
 KW anti-arthritic; thrombolytic; antirheumatic; antipruritic;
 KW immunosuppressant; pancreatitis; deep vein thrombosis; psoriasis;
 KW rheumatoid arthritis; myocardial infarction; transplant rejection.

XX Homo sapiens.

OS Synthetic.

PN WO963090-A2.

PD 09-DEC-1999.

PF 03-JUN-1999; 99WO-US12276.

PR 03-JUN-1998; 98US-0087885.

XX (SCIO-) SCIOS INC.

PI White RT, Damm D, Lesikar DD, McFadden K, Garrick BL, Lucas AB;
 PI Pollitt NS, Lam AO;

DR WPI; 2000-105699/09.

PT Novel enzyme inhibitors especially used to reduce postoperative,
 PT bleeding

PS Example 4; Fig 64; 151pp; English.
 XX
 CC The present invention describes protease inhibitors that
 CC are analogues of the kunitz protease inhibitor (KPI) domain of the
 CC amyloid precursor protein. The protease inhibitors can be used used to
 CC treat or prevent disorders associated with increased activity of serine
 CC proteases, specifically blood loss during surgery (particularly
 CC cardiopulmonary bypass surgery where plasma proteases are activated by
 CC contact with surfaces in the heart-lung machine), but also other
 CC conditions such as pancreatitis; deep vein thrombosis; rheumatoid
 CC arthritis; psoriasis; myocardial infarction; and transplant rejection.
 CC They are also for organ preservation and to promote wound healing. In
 CC vitro the protease inhibitors may be used to inhibit serine proteases
 CC during preparation of cell extracts. The protease inhibitors are based
 CC on a human peptide sequence so are unlikely to be immunogenic, can be
 CC produced at high levels in recombinant expression systems, and can
 CC inhibit a wide range of serine proteases. They are more potent or
 CC specific than known inhibitors. The present sequence represents a KPI
 CC variant which is given in an example from the present invention.
 XX
 SQ Sequence 61 AA;
 Query Match 31.5%; Score 99; DB 21; Length 61;
 Best Local Similarity 39.6%; Pred. No. 2.3e-05;
 Matches 19; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
 Oy 6 QACGTPSHLVLMHYDPQRCGCMTPARCGDGAARGFETYEACQAC 53
 Db 10 qaelgpcramlaryfdvlegkcapfigygcgdnrnmidteeycmavc 57
 RESULT 8
 AAR78563
 ID AAR78563 standard; peptide: 58 AA.
 AC AAR78563;
 XX
 DT 01-MAR-1996 (first entry)
 XX
 DE Human ITI-KUDOM 2 derivative DPI-4.2.2.
 XX
 KW Human; lipoprotein-associated coagulation inhibitor; peptide library;
 KW inhibitor; plasmin; bovine; pancreatic trypsin inhibitor; kunitz domain;
 KW fibrinolysis; fibrinogenolysis; bleeding; thrombolytic.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 11 /note- "residue change: Arg to Thr"
 FT MISC-difference 17 /note- "residue change: Phe to Arg"
 FT MISC-difference 18 /note- "residue change: Ile to Phe"
 FT
 PN WO9518830-A2.
 PN
 PD 13-JUL-1995.
 PD
 PE 11-JAN-1995; 95WO-US00298.
 PE
 PR 10-MAR-1994; 94US-0208265.
 PR 11-JAN-1994; 94US-0179658.
 XX
 PA (PROT-) PROTEIN ENG CORP.
 XX
 PI Ladner RC, Markland W;
 XX
 DR WPI; 1995-255042/33.
 DR
 XX Novel plasmin inhibiting protein comprising a Kunitz Domain - useful
 PT to prevent/treat disorders attributable to excess plasmin activity.

XX
 PS Claim 3; Page 35; 59pp; English.
 XX
 CC The peptides AAR78435-78570 are derivatives of the kunitz domains from
 CC a variety of plasmin inhibitors e.g. the human lipoprotein-associated
 CC coagulation inhibitor (LACI) Kunitz domains 1, 2 or 3. The peptides
 CC were designed based on the kunitz domains and are named Designed Plasmid
 CC Inhibitor (DPI). This peptide is designed based on the human ITI Kunitz
 CC domain 2 (Kudom) (AAR78555).
 CC The peptides can be used to prevent or treat a clinical condition
 CC exacerbated by plasmin e.g. inappropriate fibrinolysis or
 CC fibrinogenolysis; excessive bleeding associated with thrombolytics.
 XX
 SQ Sequence 58 AA;
 Query Match 31.4%; Score 98.5; DB 16; Length 58;
 Best Local Similarity 35.2%; Pred. No. 2.6e-05;
 Matches 19; Conservative 7; Mismatches 23; Indels 5; Gaps 1;
 Oy 5 VQAC-----TGPTSPHVLVLMHYDPQRCGCMTPARCGDGAARGFETYEACQAC 53
 Db 2 vaacnlpivtgsprarfqlwaidavkvcvlfpygscgngnkfysekcreyc 55
 RESULT 9
 AAR36764
 ID AAR36764 standard; Protein; 101 AA.
 AC AAR36764;
 XX
 DT 16-SEP-1993 (first entry)
 XX
 DE Factor X inhibiting peptide, R7E/Q15K/Y42E, coding sequence.
 XX
 KW Factor Xa; inhibition; urinary trypsin inhibitor; Bikunin; elastase;
 KW substitution; mutation; secretion; drug; URI; infestation; shock;
 KW pancreatitis; ischaemic heart disease; rheumatoid arthritis.
 XX
 OS Escherichia coli, strain JES505 (pW738).
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 42 /label- R7E
 FT MISC-difference 50 /label- Q15K
 FT MISC-difference 77 /label- Q15K
 FT Protein /label- Y42E
 FT 24..74 /note- "may opt. be sequentially extended at
 FT the N- and C-terminal by one amino acid to
 FT include bases upto position 34 and upto
 FT position 101 respectively, may also be
 FT extended at the N-terminal by the sequence
 FT given in AAR39758"
 XX
 PN EP543240-A.
 PN
 PD 26-MAY-1993.
 PD
 PE 06-NOV-1992; 92EP-0119083.
 PE
 PR 08-NOV-1991; 91JP-0293472.
 PR 12-MAY-1992; 92JP-0119289.
 XX
 PA (MOCH) MOCHIDA PHARM CO LTD.
 XX
 PI Kanamori T, Morishita H, Nobuhara M;
 XX
 DR WPI; 1993-168945/21.
 DR N-PSDB; AAQ41882.
 XX
 PT New polypeptide inhibiting protease(s), esp. FXa - used for

PT treating multiple organ failure, shock, pancreatitis,
PT disseminated intravascular coagulation, etc.

PS Disclosure; Page 71; 130pp; English.

XX The sequences given in AAR36751-67 represent polypeptides which have
CC inhibitors based on human kunitz protease inhibitor (KPI). Protease
CC inhibitors such as these, can bind to and inhibit the activity of
CC serine proteases. They can be used to treat conditions associated with
CC increased activity of serine proteases, e.g. cardiopulmonary bypass
CC (CPB)-induced pulmonary injury, post-CPB pulmonary injury, pancreatitis,
CC allergy-induced protease release, deep vein thrombosis,
CC thrombocytopenia, rheumatoid arthritis, adult respiratory distress
CC syndrome, chronic inflammatory bowel disease, psoriasis,
CC hyperfibrinolytic haemorrhage and myocardial infarction. They can also be
CC used for organ preservation and wound healing. They can also be used for
CC the inhibition of serine protease activity in vitro, e.g. during the
CC preparation of cellular extracts to prevent degradation of cellular
CC proteins. The protease inhibitors show more potent and specific serine
CC protease inhibition activity than known serine protease inhibitors.

CC Sequence 101 AA;

Query Match 31.4%; Score 98.5; DB 14; Length 101;
Best Local Similarity 36.5%; Pred. No. 4.6e-05;
Matches 19; Conservative 8; Mismatches 22; Indels 3; Gaps 1;

OY 2 LPPVQACTGPTSPHLVLMHYDPORGCMTPPARCGDGAARGFETYYEACQAC 53
Db 38 lplve--gpcrafiklwaidavkvcvlfpygqcgngkfeseketeyc 86

RESULT 10

AAM18397
ID AAM18397 standard; peptide; 61 AA.

AC AAM18397;

DT 06-AUG-1997 (first entry)

DE KPI(-4 to 57) N39A.

KW Kunitz protease inhibitor; KPI; serine protease; rheumatoid arthritis;
KW cardiopulmonary bypass; CPB-induced pulmonary injury; thrombocytopenia;
KW post-CPB pulmonary injury; pancreatitis; inhibition; psoriasis; CPB;
KW allergy-induced protease release; adult respiratory distress syndrome;
KW hyperfibrinolytic haemorrhage; wound healing; myocardial infarction;
KW chronic inflammatory bowel disease; organ preservation; human;
KW deep vein thrombosis.

OS Homo sapiens.
OS Synthetic.

PN WO9635788-A2.

PD 14-NOV-1996.

PF 08-MAY-1996; 96WO-US06384.

PR 06-MAY-1996; 96US-0643731.

PR 08-MAY-1995; 95US-0436555.

PA (SCIO-) SCIOS INC.

PI Damm D, Garrick BL, Lesikar DD, McFadden K, White TR;

DR WPI; 1996-518676/51.

PT New protease inhibitor peptide based on Kunitz Protease Inhibitor -
PT used for treating e.g. cardio:pulmonary bypass-induced pulmonary
PT injury, pancreatitis, allergy-induced protease release, deep vein
PT thrombosis, etc

PS Claim 9; Page -; 138pp; English.

XX The sequences given in AAM18395-431 represent mutant protease
CC inhibitors based on human kunitz protease inhibitor (KPI). Protease
CC inhibitors such as these, can bind to and inhibit the activity of
CC serine proteases. They can be used to treat conditions associated with
CC increased activity of serine proteases, e.g. cardiopulmonary bypass
CC (CPB)-induced pulmonary injury, post-CPB pulmonary injury, pancreatitis,
CC allergy-induced protease release, deep vein thrombosis,
CC thrombocytopenia, rheumatoid arthritis, adult respiratory distress
CC syndrome, chronic inflammatory bowel disease, psoriasis,
CC hyperfibrinolytic haemorrhage and myocardial infarction. They can also be
CC used for organ preservation and wound healing. They can also be used for
CC the inhibition of serine protease activity in vitro, e.g. during the
CC preparation of cellular extracts to prevent degradation of cellular
CC proteins. The protease inhibitors show more potent and specific serine
CC protease inhibition activity than known serine protease inhibitors.

XX Sequence 61 AA;

Query Match 31.2%; Score 98; DB 17; Length 61;
Best Local Similarity 39.6%; Pred. No. 3.1e-05;
Matches 19; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

OY 6 QACTGPTSPHLVLMHYDPORGCMTPPARCGDGAARGFETYYEACQAC 53
Db 10 qaeqtporamlsrwyfdvteqkcapffygqcggarndfdeeycmavc 57

RESULT 11

AAM18401
ID AAM18401 standard; peptide; 61 AA.

AC AAM18401;

DT 06-AUG-1997 (first entry)

DE KPI(-4 to 57) M15A, S17Y, R18H, N39A.

KW Kunitz protease inhibitor; KPI; serine protease; rheumatoid arthritis;
KW cardiopulmonary bypass; CPB-induced pulmonary injury; thrombocytopenia;
KW post-CPB pulmonary injury; pancreatitis; inhibition; psoriasis; CPB;
KW allergy-induced protease release; adult respiratory distress syndrome;
KW hyperfibrinolytic haemorrhage; wound healing; myocardial infarction;
KW chronic inflammatory bowel disease; organ preservation; human;
KW deep vein thrombosis.

OS Homo sapiens.
OS Synthetic.

PN WO9635788-A2.

PD 14-NOV-1996.

PF 08-MAY-1996; 96WO-US06384.

PR 06-MAY-1996; 96US-0643731.

PR 08-MAY-1995; 95US-0436555.

PA (SCIO-) SCIOS INC.

PI Damm D, Garrick BL, Lesikar DD, McFadden K, White TR;

DR WPI; 1996-518676/51.

PT New protease inhibitor peptide based on Kunitz Protease Inhibitor -
PT used for treating e.g. cardio:pulmonary bypass-induced pulmonary
PT injury, pancreatitis, allergy-induced protease release, deep vein
PT thrombosis, etc

PS Example; Page -; 138pp; English.

XX The sequences given in AAM18395-431 represent mutant protease

SQ Sequence 61 AA;

```

Oy      6 QACTGPTSPHLVLWVHYDPQGGCMTFPARCGDGAARGFETYEACQAC 53
          || ||| : ||| : | | | | | | | | | |
Db     10 qactgpcraaalyhwyfdvtgkcapffygscgarnfdteeycmavc 57

```

```

RESULT 12
AAY68092
ID AAY68092 standard; Protein; 61 AA.

```

DT 13-APR-2000 (first entry)

DE Kunitz protease inhibitor variant DD128.

KM knittz protease inhibitor; KPI: serine protease; kallikrein; plasmin
KM anyold precursor protein; coagulation factor; blood loss; caudant;
KM cardiopulmonary bypass surgery; anticoagulant; anti-inflammatory;
KM anti-arthritic; thrombolytic; antithematic; antipsoriatic;
KM immunosuppressant; pancreatitis; deep vein thrombosis; psoriasis;
KM rheumatoid arthritis; myocardial infarction; transplant rejection.

OS	Homo sapiens.
OS	Synthetic.

PN WO9963090-A2.

PD 09-DEC-1999.

PF 03-JUN-1999; 99WO-US12276.

PR 03-JUN-1998; 98US-0087885.

PA (SCIO-) SCIOS INC.
XX

PI White RT, Damm D, Lesikar DD, McFadden K, Garrick BL, Lucas AB,
PI Pollitt NS, Lam AO;

DR WPI; 2000-105699/09.

PT	Novel enzyme inhibitors especially used to reduce postoperative bleeding
PT	-

PS Example 4; Flg 64; 151pp; English.

CC The present invention describes protease inhibitors that
CC are analogues of the kunitz protease inhibitor (KPI) domain of the
CC amyloid precursor protein. The protease inhibitors can be used to
CC treat or prevent disorders associated with increased activity of serine

proteases, specifically blood loss during surgery (particularly cardiopulmonary bypass surgery where plasma proteases are activated by contact with surfaces in the heart-lung machine), but also other conditions such as pancreatitis; deep vein thrombosis; rheumatoid arthritis; psoriasis; myocardial infarction; and transplant rejection. They are also for organ preservation and to promote wound healing. In vitro the protease inhibitors may be used to inhibit serine proteases during preparation of cell extracts. The protease inhibitors are based on a human peptide sequence so are unlikely to be immunogenic, can be produced at high levels in recombinant expression systems, and can inhibit a wide range of serine proteases. They are more potent or specific than known inhibitors. The present sequence represents a KPI variant which is given in an example from the present invention.

50 Sequence 61 AA;

Query Match	31.28	Score 98	DB 21	Length 61
Best Local Similarity	39.68	Pred. No.	3.1e-05	
Matches 19, Conservative	4	Mismatches	25	Indels 0; Gaps 0

QY 6 QACTGPTSPHLVLMHYDPQRGGCMTFPARCDDGARGFETYEACQAC 53
 |||| :||:| | | | | | | |
 Db 10 qaeltpcramislwYfdvteqkcapffygcygnrnfdtteycmavc 57

RESULT 13	
AAV68115	
ID AAV68115	standard; Protein; 61 AA

DT 13-APR-2000 (first entry)

DE Kunitz protease inhibitor variant TW6148
 117

KM amyloid precursor protein; coagulation factor; blood loss; cardiac;
KM cardiopulmonary bypass surgery; anticoagulant; anti-inflammatory;
KM anti-arthritic; thrombotic; antirheumatic; antispasmodic;
KM immunosuppressant; pancreaticitis; deep vein thrombosis; psoriasis;
KM rheumatoid arthritis; myocardial infarction; transplant rejection.

OS Homo sapiens.
OS Synthetic.

PN W09963090-A2.
XX

XX

XX

XX

XX

PI Pollitt NS, Lam AO;

DR WPI; 2000-105699/09
XX

PT	bleeding	-

PS Example 4; Fig 64; 151pp; English.
XX

CC are analogues of the Kunitz protease inhibitor (KPI) domain

CC treat or prevent disorders associated with increased activity of serine
CC proteases, specifically blood loss during surgery (particularly
CC cardiopulmonary bypass surgery where plasma proteases are activated by
CC contact with surfaces in the heart-lung machine), but also other

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:22 ; Search time 144.8 Seconds

(without alignments)
47.478 Million cell updates/sec

Title: US-09-819-136-2_COPY_412_458

Perfect score: 239

Sequence: 1 PRTPPCRCRKLRSKLALSLC.....GRLTEVLEPEPPAGIARVA 47

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

Listing first 45 summaries

```

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organella:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	24.7	742	11	09GCK5 mus musculu
2	57.5	24.1	293	4	09UHW9
3	57	23.8	276	2	046612
4	56	23.4	3161	2	054511
5	56	23.0	158	4	09BVL3
6	55	23.0	233	4	09BVL3
7	55	23.0	253	4	09BVL3
8	55	23.0	420	4	09BVL3
9	55	23.0	429	4	09BVL3
10	54.5	22.8	249	2	09LJH0
11	54.5	22.6	446	1	058932
12	54	22.6	770	10	049846
13	53.5	22.4	196	2	09EVL4
14	53.5	22.4	196	2	09EVL3
15	53.5	22.4	196	2	09EVL4
16	53.5	22.4	446	10	09F6N4
17	53	22.2	132	2	09R2S1
18	53	22.2	147	2	055833
19	53	22.2	191	1	050562

20	53	22.2	266	10	09F293	09F293 arabidopsis
21	53	22.2	363	1	027188	027188 methanobact
22	53	22.2	378	10	023227	023227 arabidopsis
23	53	22.2	532	5	09V7B3	09V7B3 drosophila
24	52.5	22.0	302	5	09VXK0	09VXK0 drosophila
25	52.5	22.0	380	2	052289	052289 agrobacteri
26	52	21.8	295	11	035297	035297 mus musculu
27	52	21.8	295	11	008862	008862 mus musculu
28	52	21.8	295	11	P97299	P97299 mus musculu
29	52	21.8	306	2	09UXY6	09UXY6 neisseria m
30	52	21.8	483	2	09HY93	09HY93 pseudomonas
31	52	21.8	1109	3	09UV93	09UV93 trichophyto
32	52	21.8	2352	2	09HYR8	09HYR8 pseudomonas
33	51	21.3	283	6	09B8G6	09B8G6 oryctolagus
34	51	21.3	291	5	P91386	P91386 caenorhabdi
35	51	21.3	1164	5	09N6X6	09N6X6 trypanosoma
36	50.5	21.1	144	11	09QY87	09QY87 rattus norv
37	50.5	21.1	168	10	09Z595	09Z595 arabidopsis
38	50.5	21.1	214	1	049603	049603 methanopyru
39	50.5	21.1	220	1	09Y9C7	09Y9C7 aeropyrum p
40	50.5	21.1	297	5	09UAX9	09UAX9 crithidia f
41	50.5	21.1	317	5	09VTV2	09VTV2 drosophila
42	50.5	21.1	421	10	09SY90	09SY90 arabidopsis
43	50.5	21.1	475	2	09X9M5	09X9M5 streptomyce
44	50.5	21.1	479	10	09FGN3	09FGN3 arabidopsis
45	50.5	21.1	979	2	053609	053609 mycobacteri

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	742 AA.
09GCK5	09GCK5			
AC	01-JUN-2001 (T-EMBLrel. 17, Created)			
DT	01-JUN-2001 (T-EMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (T-EMBLrel. 17, Last annotation update)			
DE	2410024C1SRK PROTEIN.			
GN	2410024C1SRK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Bazzin G.,			
RA	Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guinacich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyo-oka K., Wang K.H., Wetz C., Whitlaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohatsu S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
DR	EMBL AK010578; BAB27041.1; -			
DR	MED: MG1:1919438; 2410024C1SRK.			
SQ	SEQUENCE 742 AA; 84086 MW; CFC44D34D96ACBA7 CRC64;			

Query Match 24.7%; Score 59; DB 11; Length 742;
 Best Local Similarity 39.4%; Pred. No. 23;
 Matches 13; Conservative 4; Mismatches 12; Indels 4; Gaps 1;

OY 1 PTPPCACRRLRSKLTALISCRSDFAVGRLEVEPEPAGGIARY 33
 DB 709 PRNSPCLTTRAROSLSKSLCOTE---GRCTRM 737

RESULT 2

O9UJH9 PRELIMINARY; PRT; 293 AA.

AC O9UJH9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-JUN-2001 (TREMBLrel. 13, last sequence update)
 DE C380A1.2.1 (NOVEL PROTEIN (ISOFORM 1)) (UNKNOWN) (PROTEIN FOR
 DE MGC:2601).
 GN C380A1.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wallis J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: 297653; CAB56188.1;
 DR EMBL: BC000662; AAH00662.1;
 SQ SEQUENCE 293 AA; 31207 MW; 45B9136BF9153F28 CRC64;

Query Match 24.1%; Score 57.5; DB 4; Length 293;
 Best Local Similarity 30.4%; Pred. No. 14;
 Matches 17; Conservative 6; Mismatches 22; Indels 11; Gaps 2;

OY 1 PTPPP-----CRACRLRSKLTALISCRSDFAVGRLEVEPEPAGGIARY 46
 DB 156 PELPQANGLSVGDGACRCRS-DAEILLACISDFVHGHVTHDELQESVTV 210

RESULT 3

O46612 PRELIMINARY; PRT; 276 AA.

AC O46612;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE IS 1222 GENE ORF-A AND ORF-B.
 OS Enterobacter agglomerans.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pantosa.
 OX NCB1_TaxID=549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KLEBERGER, 1983;
 RC MEDLINE=9525564; PubMed=7737514;
 RA Steidl H.D., Lewecke P.M.;
 RT "ISI222: analysis and distribution of a new insertion sequence in
 RT Enterobacter agglomerans 339.";
 RL Gene 156:37-42(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Steidl H.D., Siddavattam D.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96422758; PubMed=8825375;

RA Steidl H.D., Siddavattam D., Klingmüller W.;
 RT "Self-transmissible nif plasmid (pEA) of Enterobacter agglomerans
 RT 339: molecular cloning and evidence for the existence of similar nif
 RT clusters on dissimilar plasmids in Enterobacter strains.";
 RL Plasmid 34:223-228(1995).
 RN [4]
 RP SEQUENCE FROM N.A.

RA Steidl H.D.;
 RL Thesis (1995), Universitaet Bayreuth, Germany.
 DR EMBL: X78052; CAA54973.1;
 DR EMBL: X81893; CAA57482.1;
 DR InterPro: IPR001584; Rve.
 DR Pfam: PF00665; rve.1
 SQ SEQUENCE 276 AA; 31718 MW; 0B72505B53122509 CRC64;

Query Match 23.8%; Score 57; DB 2; Length 276;
 Best Local Similarity 40.5%; Pred. No. 15;
 Matches 17; Conservative 6; Mismatches 11; Indels 8; Gaps 2;

OY 7 RACRLRSKLTALISCR-----SDFAVGRLEVEPEPAG 41
 DB 13 RACRL-TGLISTCKYEAQRPADAHLSGRITELALERRRG 53

RESULT 4

O54511 PRELIMINARY; PRT; 3161 AA.

AC O54511;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE HMP1 PROTEIN.
 GN IRP1.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCB1_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-WA-314;
 RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RL EMBL: Y12527; CAA73127.1;
 DR InterPro: IPR001227; Acyltransf_domain.
 DR InterPro: IPR001525; C5_DNA_meth.
 DR InterPro: IPR001242; DUF4.
 DR InterPro: IPR000794; ketoacyl-synt.
 DR InterPro: IPR001601; Meth-transf.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR InterPro: IPR000051; SAM_bind.
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00698; Acyl_transf.1.
 DR Pfam: PF00668; Condensation.1.
 DR Pfam: PF00109; ketoacyl-synt.1.
 DR Pfam: PF00550; pp-binding.2.
 DR Pfam: PF00975; Thioesterase.1.
 DR PROSITE: PS0075; ACP_DOMAIN.2.
 DR PROSITE: PS00606; B_KETOACYL_SYNTHASE.1.
 DR PROSITE: PS00095; C5_MTAASE.2; UNKNOWN.1.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN.2.
 KW Phosphopantetheine; Phosphatase.
 SQ SEQUENCE 3161 AA; 348654 MW; 84E07081A19EBD95 CRC64;

RP SEQUENCE FROM N.A.
 RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Oobayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isonaga T., Sugano S.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK024644; BAB14944.1; -
 SO SEQUENCE 253 AA; 27348 MW; 3211AEF55AC426BA CRC64;

Query Match 23.0%; Score 55; DB 4; Length 253;
 Best Local Similarity 34.9%; Pred. No. 25;
 Matches 15; Conservative 9; Mismatches 15; Indels 4; Gaps 2;

QY 5 PCRACRLR---SKLALSLCRSDPAIVGRLTVELEPEPAAGGTA 44
 Db 153 PQAWEIOVSSSLA-DICRAHNAVGMQTVTEQAAGSSA 194

RESULT 9
 Q9BTP9 PRELIMINARY; PRT; 420 AA.
 AC Q9BTP9;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MGC:10791).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDOMETRIAL ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC003533; AA003533.1; -
 SQ SEQUENCE 420 AA; 44402 MW; 75C4C479C67BA0D6E CRC64;

Query Match 23.0%; Score 55; DB 4; Length 420;
 Best Local Similarity 36.5%; Pred. No. 42;
 Matches 19; Conservative 5; Mismatches 20; Indels 8; Gaps 2;

QY 2 RTPPCRRLSKLALSLCRSD-----FAIVGRLTVELEPEPAAGGTA 45
 Db 290 RPPGSGRLRLGTALLLFRSDARGHAGGFALTGRLDADADPRAPRESSAQ 341

RESULT 10
 Q9L3H0 PRELIMINARY; PRT; 249 AA.
 AC Q9L3H0;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHEICAL 28.0 KDA PROTEIN (FRAGMENT).
 GN REPC.
 OS Rhizobium leguminosarum (biovar viciae).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 CX NCBI_TaxID=387;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-F1170;
 RA Palmer K.M., Turner S.L., Young J.P.W.;
 RT "Sequence diversity of the plasmid replication gene repC in the
 RL Rhizobiaceae."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ276304; CAB77030.1; -
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 249 249

SQ SEQUENCE 249 AA; 28034 MW; DBEC76884593C010 CRC64;

Query Match 22.8%; Score 54.5; DB 2; Length 249;
 Best Local Similarity 45.9%; Pred. No. 29;
 Matches 17; Conservative 5; Mismatches 10; Indels 5; Gaps 2;

QY 10 RLRSKLALSLCRSDPAIVGRLTVELEPEPAAGGTA 46
 Db 157 RLRR--LTLCRRD--VAKILEMALEEGAGADWSRV 188

RESULT 11
 O58932 PRELIMINARY; PRT; 446 AA.
 AC O58932;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE HYPOTHEICAL 51.7 KDA PROTEIN PH1251.
 GN PH1251.
 OS Pyrococcus horikoshii.
 CC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 CX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O13;
 RX MEDLINE=96344137; PubMed=9679194;
 RA Kawarayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masachi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000005; BAA30352.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 446 AA; 51659 MW; 3EDC511D4D3BC15 CRC64;

Query Match 22.8%; Score 54.5; DB 1; Length 446;
 Best Local Similarity 27.9%; Pred. No. 52;
 Matches 12; Conservative 7; Mismatches 13; Indels 11; Gaps 1;

QY 1 PRTPCRRLSKLALSLCRSDPAIVGRLTVELEPEPAAGGTA 43
 Db 95 PSTPLCEICGRHRSRSPYRSEFPLIG-----SGGV 126

RESULT 12
 O49846 PRELIMINARY; PRT; 770 AA.
 AC O49846;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE INWARD POTASSIUM CHANNEL ALPHA SUBUNIT (FRAGMENT).
 OS Egeria densa.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Hydrocharitaceae; Egeria.
 CX NCBI_TaxID=55453;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LEAF;
 RA May M., Saalbach G., Dahse I.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ225805; CAA12645.1; -
 DR HSSP: P55273; 1B18.
 DR Mendel: 27767; Bgde:1146; 27767.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000636; Ank_chan_non_1lig.

OG Plasmid uncharacterised.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:41 ; Search time 45.78 Seconds

(without alignments)
37.642 Million cell updates/sec

Title: US-09-819-136-2_COPY_412_458

Perfect score: 239
Sequence: 1 PRTPRCGRCLRLSKLALSLC.....GRLPVELEPEAGGIARVA 47

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: SWISSProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	25.5	1184	1	POL2_GFLV
2	56.5	23.6	423	1	YD94_METTH
3	53	22.2	188	1	TRAF_RHSN
4	53	22.2	941	1	CHRD_XENLA
5	52.5	22.0	159	1	Y707_TREPA
6	52	21.8	397	1	PGK_XANFL
7	51.5	21.5	226	1	VATE_HETSC
8	51.5	21.5	231	1	NEFL_PIG
9	51.5	21.5	506	1	NEFC_PIG
10	51.5	21.5	996	1	PROL_SARPE
11	51	21.3	381	1	PBP4_NOCLA
12	51	21.3	664	1	UL47_HSV1F
13	51	21.3	1042	1	CORI_HUMAN
14	50.5	21.1	431	1	NEFC_CHICK
15	50.5	21.1	439	1	NEFC_MOUSE
16	50.5	21.1	449	1	PCOI_HUMAN
17	50	20.9	1634	1	PCOI_HUMAN
18	49.5	20.7	373	1	EGON_DROME
19	49	20.5	214	1	DEVS_MYXXA
20	48.5	20.3	126	1	ANPC_HUMAN
21	48.5	20.3	295	1	SP19_YEAST
22	48.5	20.3	354	1	VATC_HORVU
23	48.5	20.3	364	1	WNT6_MOUSE
24	48.5	20.3	365	1	WNT6_HUMAN
25	48.5	20.3	393	1	IRK9_HUMAN
26	48.5	20.3	439	1	NEFC_CHICK
27	48.5	20.3	493	1	DCMC_HUMAN
28	48.5	20.3	508	1	NEFC_HUMAN
29	48.5	20.3	569	1	UVRA_VITST
30	48.5	20.3	614	1	DDX5_HUMAN
31	48.5	20.3	614	1	DDX5_MOUSE
32	48.5	20.3	663	1	PDIS_HUMAN
33	48.5	20.3	950	1	UVRA_NETGO

ALIGNMENTS

34	48.5	20.3	1124	1	TRPL_DROME	P48994 drosophila
35	48	20.1	108	1	Y187_BUCAI	P57284 buchera ap
36	48	20.1	374	1	ADH1_APPAU	P49665 apteryx aus
37	48	20.1	473	1	ADPB_BACCA	P41009 bacillus ca
38	48	20.1	473	1	ATPB_BACCP3	P07677 bacillus ps
39	48	20.1	473	1	ATPB_BACST	P42006 bacillus st
40	48	20.1	594	1	SYD_MYCVU	O50649 mycobacteri
41	48	20.1	797	1	SYM_ARATH	O95V55 arabidopsis
42	47.5	19.9	316	1	OCDS_HUMAN	O9UG57 homo sapien
43	47.5	19.9	512	1	PERI_VOICA	P81131 volvox cart
44	47.5	19.9	861	1	ORCI_HUMAN	O13415 homo sapien
45	47.5	19.9	954	1	MKA_HUMAN	O02779 homo sapien

RESULT 1
POL2_GFLV STANDARD; PRT; 1184 AA.
AC P18474; O83857;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA2 POLYPROTEIN (131 KDA PROTEIN) [CONTAINS: COAT PROTEIN; 75 KDA DE PROTEIN].
OS Grapevine fanleaf virus (GFLV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
NC Nepovirus.
OX NCBI_TaxID=12274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL3;
RX MEDLINE=90324929; PubMed=2374004;
RA Serghini M.A., Fuchs M., Plinck M., Reinbolt J., Walter B., Plinck L.;
RT "RNA2 of grapevine fanleaf virus: sequence analysis and coat protein RT cistron location".
RL J. Gen. Virol. 71:1433-1441(1990).
CC -!- FUNCTION: THE PROTEIN LOCATED AT THE N-TERMINAL END OF THE NEPOVIRUS POLYPROTEIN COULD BE REQUIRED FOR NEMATODE TRANSMISSION OF THE VIRUS.
CC -!- PTM: THE RNA2 POLYPROTEIN IS CLEAVED BY AN RNA1-ENCODED PROTEASE TO YIELD THE MATURE COAT PROTEIN AND AN 66 KDA PROTEIN WHICH IS FURTHER CLEAVED INTO TWO PRODUCTS OF APPROXIMATELY 46 AND 20 KDA.
CC -!- SIMILARITY: TO THE RNA2 POLYPROTEIN OF OTHER NEPOVIRUSES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X15907; CAA34779.1; ALT-INIT.
DR EMBL; X15907; CAA34780.1; ALT-INIT.
KW Polypeptide; Coat protein.
FT DOMAIN 1 630 75 KDA PROTEIN.
FT DOMAIN 681 1184 COAT PROTEIN.
SQ SEQUENCE 1184 AA; 131608 MW; E206B2AASB0675D0 CRC64;

Query Match 25.5%; Score 61; DB 1; Length 1184;
Best Local Similarity 41.7%; Pred. No. 3;
Matches 20; Conservative 5; Mismatches 21; Indels 2; Gaps 1;

QY 1 PRTPRCGRCLRLSKLALSLCSDP--ALVGRLEVELEPEAGGIARV 46
DB 184 PTPPPAPYFRAVGAFAPTSEFVAIVETRLRESRAALFAEL 231

RESULT 2
YD94_METTH

```

0X NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed-9163424;
RA Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A.,
RT Perret X.;
    "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -I- FUNCTION: INVOLVED IN CONJUGAL TRANSFER OF THE PLASMID
    (POTENTIAL).
CC -I- SUBCELLULAR LOCATION: PERIPLASMIC (PROBABLE).
CC -I- SIMILARITY: STRONG, TO A-TWEEFACTINS TI PLASMID TRAF. ALSO TO
    E-COLI INCP PLASMIDS TRAF.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation .
CC The European Bioinformatics institute. There are no restrictions on its use
CC by non-profit institutions as long as its content is in no way used for
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
or send an email to license@isb-slb.ch).
-----
CC EMBL: AE000069; AAB91647.1; -.
KW Conjugation; Plasmid; Periplasmic; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 188 PROBABLE CONJUGAL TRANSFER PROTEIN TRAF.
SQ SEQUENCE 188 AA; 19999 MW; 93A40D18C89AC69 CRC64;
-----
Oy Query Match 22.2%; Score 53; DB 1; Length 188;
Best Local Similarity 34.1%; Pred. No. 6.1;
Matches 14; Conservative 4; Mismatches 15; Indels 8; Gaps 1.;
Db 74 PRPSAMREARMNGYLKAGLCGGGIMPLIKSVLAIVAGQREV 114
1 PRPPCRACRLSKSLATLSLC-----RSDPAIVGRLEEV 33
||| | | : | | || : | | : ||
74 PRPSAMREARMNGYLKAGLCGGGIMPLIKSVLAIVAGQREV 114
-----
RESULT 4
CHRD_XENLA STANDARD; PRT; 941 AA.
ID CHRD_XENLA
AC Q91713;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CHORDIN PRECURSOR (ORGANIZER-SPECIFIC SECRETED DORSALIZING FACTOR).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Dorsal 11p;
RX MEDLINE=95094250; PubMed-800117;
RA Sasai Y., Lu B., Steinbeisser H., Geissert D., Gont L.K.,
de Robertis E.M.;
    "Xenopus chordin: a novel dorsalizing factor activated by organizer-
specific homeobox genes.";
RL Cell 79:779-790(1994).
CC -I- FUNCTION: POTENT DORSALIZING FACTOR. HAS POTENT AXIS-FORMING
ACTIVITIES INCLUDING THE ABILITY TO RECRUIT NEIGHBORING CELLS INTO
SECONDARY AXES. REGULATES CELL-CELL INTERACTIONS IN THE ORGANIZING
CENTERS OF HEAD, TRUNK AND TAIL DEVELOPMENT.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- TISSUE SPECIFICITY: PRECORDIAL PLATE, NOTOCHORD, CHORDONEURAL
HINGE. EXPRESSION CLOSELY FOLLOWS THE AREAS OF THE EMBRYO THAT
EXPRESS GOOSECOED AND XNOT2.
CC -I- DEVELOPMENTAL STAGE: EMBRYOGENESES.
CC -I- SIMILARITY: BELONGS TO THE CHORDIN FAMILY.
CC -I- SIMILARITY: CONTAINS 4 WWFC DOMAINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35764; AAC4222.1; -
DR InterPro: IPR001007; VMEC.
DR Pfam: PF00093; VWC; 4.
DR SMART: SM00214; VWC; 4.
DR PROSITE: PS01208; VMEC; 2.
KM Developmental protein; Repeat: Glycoprotein; signal.
FT STGNL 1 19 POTENTIAL.
FT CHAIN 20 941 CHORDIN.
FT DOMAIN 43 116 VMEC 1.
FT DOMAIN 693 750 VMEC 2.
FT DOMAIN 769 837 VMEC 3.
FT DOMAIN 857 919 VMEC 4.
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 941 AA; 104947 MW; 7DIC3262064C7323 CRC64;

Query Match 22.28; Score 53; DB 1; Length 941;
Best Local Similarity 32.58; Pred. No. 28;
Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

OY 6 CRACRLRSKLALSLCRSDFAIVGRLTVELEPEAAGIAR 45
DB 237 CGTWSLRSTLRRLRMGHILVSLVTTLSPEISGKIIVK 276

RESULT 5
Y707_TREPA
ID Y707_TREPA STANDARD; PRT; 159 AA.
AC 083705;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHEICAL PROTEIN TP0707.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_Taxid:160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RX MEDLINE:98332770; PubMed:9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Kralak H., Richardson D., Howell J.K., Chidambaram M., Ullrich T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001244; AAC65694.1; -
DR TIGR: TP0707; -
KM Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 91 113 POTENTIAL.
SQ SEQUENCE 159 AA; 16909 MW; DC363F323456F1CC CRC64;

Query Match 22.08; Score 52.5; DB 1; Length 159;
Best Local Similarity 40.08; Pred. No. 6.1;
Matches 18; Conservative 5; Mismatches 19; Indels 3; Gaps 2;

OY 5 PCRACRLRSKLALSLCRSDFAIVGRLTVELEPEAAGIAR 47
DB 14 PCSACTC-TQYAVQALSMRSGMRLKRCVRRRRCALMAIVRA 57

RESULT 6
PGK_XANFL
ID PGK_XANFL STANDARD; PRT; 397 AA.
AC P50314;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN PGK.
OS Xanthobacter flavus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hypnomicrobium group; Xanthobacter.
OX NCBI_Taxid:281;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H4-14;
RX MEDLINE:95014048; PubMed:7928974;
RA Meijer W.G.;
RT "The Calvin cycle enzyme phosphoglycerate kinase of Xanthobacter
RT flavus required for autotrophic CO2 fixation is not encoded by the
RT cbb operon.";
RL J. Bacteriol. 176:6120-6126(1994).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE - ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOSOL.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08462; AAA62185.1; -
DR EMBL: U33064; AAA96748.1; -
DR HSSP: P36204; IYPE.
DR InterPro: IPR001576; PGK.
DR Pfam: PF00162; PGK; 1.
DR PRINTS: PR00477; PGGLYCKINASE.
DR PROSITE: PS00111; PGLYCERATE_KINASE; 1.
KM Transferase; kinase; glycolysis.
SQ SEQUENCE 397 AA; 41726 MW; 2492392F0AEBDBP8 CRC64;

Query Match 21.88; Score 52; DB 1; Length 397;
Best Local Similarity 41.98; Pred. No. 17;
Matches 13; Conservative 5; Mismatches 11; Indels 2; Gaps 1;

OY 11 LRSKLALSLCRSDFAIVGRLTVELEPEAAG 41
DB 230 LGRKVGKSLCEHDLANTAR--DILKKAAG 258

RESULT 7
VATE_HFTSC

```

ID VATE_HERTSC STANDARD: PRT: 226 AA.
AC 0901G5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT E (EC 3.6.1.34) (V-ATPASE E SUBUNIT)
DE (VACUOLAR PROTON PUMP E SUBUNIT).
OS Heterodera schachtli.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
OX NCBI_TaxID=97005;
RN [1]
RP SEQUENCE FROM N.A.
RA Vercauteren I.J.R., Cheysen G.;
RT Isolation of a cDNA encoding the first vacuolar H+-ATPase subunit E
RT from plant-parasitic nematodes and its mRNA expression in pre-
RT parasitic Heterodera schachtli juveniles.
RT Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF VACUOLAR ATPASE
CC ESSENTIAL FOR ASSEMBLY OR CATALYTIC FUNCTION. V-ATPASE IS
CC RESPONSIBLE FOR ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS
CC IN EUKARYOTIC CELLS (BY SIMILARITY).
CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (COMPONENTS A TO H) ATTACHED TO
CC AN INTEGRAL MEMBRANE V0 PROTON PORE COMPLEX (COMPONENTS: A, C, C',
CC C'', AND D) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE V-ATPASE E SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ249961; CAB62552.1;
DR InterPro: IPR002842; V-ATP-synt_E.
DR Pfam: PF01991; V-ATP-synt_E.1.
DR HydroLase: ATP synthetase; Hydrogen ion transport.
KW SEQUENCE 226 AA; 26293 MW; CBB3EB2C0470C029 CRC64;
SQ

```

Query Match 21.5%; Score 51.5; DB 1; Length 226;
 Best Local Similarity 31.0%; Pred. No. 11;
 Matches 18; Conservative 6; Mismatches 13; Indels 21; Gaps 2;

```

OY 11 LRSKIALSLCRSDFAIVGR-LTEVLEE-----PEAAGGIRVA 47
DB 128 LESKVLAKCRKDEEMVARILPECLEVEYQRTGNSVYKIDNEHFLSPESAGVELLA 185

```

RESULT 8

ID	NCBI_PIG	STANDARD:	PRT:	231 AA.
AC	PI4057;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	NUCLEAR FACTOR 1 (NF-1) (CCAAT BOX-BINDING TRANSCRIPTION FACTOR) (CTF)			
DE	(TGCGA-BINDING PROTEIN) (FRAGMENT).			
GN	NFI.			
OS	Sus scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=88296883; PubMed=2841167;			
RA	Meisterernst M., Rogge L., Donath C., Gander I., Lottspeich F.,			
RA	Mertz R., Dobner T., Foeckler R., Stelzer G., Winnacker E.L.;			
RT	"Isolation and characterization of the porcine nuclear factor I (NFI)			
RT	gene."			

```

RL FEBS Lett. 236:27-32(1988).
RN [2]
RP PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=89088257; PubMed=3207762;
RA Gander I., Foeckler R., Rogge L., Meisterernst M., Schneider R.,
RA Mertz R., Lottspeich F., Winnacker E.L.;
RT "Purification methods for the sequence-specific DNA-binding protein
RT nuclear factor I (NFI) -- generation of protein sequence
RT information";
RL Biochim. Biophys. Acta 951:411-418(1988).
CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
CC TTGGCANNNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE CTF/NFI FAMILY.
CC -1- SIMILARITY: CONTAINS 1 DNA/MFI DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X12764; CAN31254.1;
DR PIR: J0463; J0463.
DR TRANSFAC: T00173;
DR TRANSFAC: T00538;
DR InterPro: IPR000647; CTF_NFI.
DR InterPro: IPR003619; Dwaftin_A.
DR Pfam: PF00859; CTF_NFI.1.
DR SMART: SM00523; DWA; 1.
DR PROSITE: PS00349; CTF_NFI; 1.
KW Transcription regulation; DNA replication; DNA-binding; Activator;
KW Nuclear protein; Multigene family; Alternative splicing.
FT DOMAIN 65 173 DWA.
FT NON_TER 231 231
SQ SEQUENCE 231 AA; 26162 MW; AA63C8BEE3D0E743 CRC64;

```

Query Match 21.5%; Score 51.5; DB 1; Length 231;
 Best Local Similarity 37.0%; Pred. No. 12;
 Matches 17; Conservative 6; Mismatches 16; Indels 7; Gaps 3;

```

OY 8 ACRSLKIALSL--CRSDF--ATVGRLTE--VLEEPFAAGIARV 46
DB 70 ASRLAKLRKDIRPCREDYFVLAITGKAPGCIVLSNPDKMKMRI 115

```

RESULT 9

ID	NCBI_PIG	STANDARD:	PRT:	506 AA.
AC	P21999;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	NUCLEAR FACTOR 1 C-TYPE (NUCLEAR FACTOR 1/C) (NFI-C) (NFI-1/C)			
DE	(CCAAT-BOX BINDING TRANSCRIPTION FACTOR) (CTF) (TGCGA-BINDING			
DE	PROTEIN).			
GN	NFI-C OR NFI.			
OS	Sus scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=90105327; PubMed=2513876;			

RA Meisterernst M., Rogge L., Foeckler R., Karaghosoff M.,
 RA Winkler E.L.;
 RT "Structural and functional organization of a porcine gene coding for
 RT nuclear factor I.";
 RL Biochemistry 28:8191-8200(1989).
 CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
 CC TTGCNNNNCCCA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
 CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
 CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-I FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 DNA/MHI DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J02875; AAA31093.1; -
 CC PIR: A33416; A33416.
 CC TRANSFAC: T00175; -
 CC TRANSFAC: T00338; -
 CC InterPro: IPR000647; CTF_NFI.
 CC InterPro: IPR003619; Dwarfin_A.
 CC Pfam: PF00859; CTF_NFI; 1.
 CC SMART: SM00523; DNA; 1.
 CC DR PROSITE: PS00349; CTF_NFI; 1.
 CC DR PROSITE: PS00347; PAR_ZN_FINGER_1; FALSE NEG.
 CC KW Transcription regulation; DNA replication; DNA-binding; Activator;
 CC Nuclear protein; Multigene family; Alternative splicing.
 CC FT DOMAIN 68 176 DNA.
 CC FT SEQUENCE 506 AA; 55461 MW; 6315A9474807F4C0 CRC64;
 CC -----
 CC Query Match 21.5%; Score 51.5; DB 1; Length 506;
 CC Best Local Similarity 37.0%; Pred. No. 24;
 CC Matches 17; Conservative 6; Mismatches 16; Indels 7; Gaps 3;
 CC QY 8 AGLRSLKSLSTL---CRSDF--ALVGRLE--VLEPEANGAGIARY 46
 CC DB 73 ASRLALRKQIRKRECEDEFLVLAITKKAPCVLSNPDQKGMRI 118
 CC -----
 CC RESULT 10
 CC PPOI_SARPE STANDARD; PRT; 996 AA.
 CC ID PPOI_SARPE
 CC AC 011208;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE POLY (ADP-RIBOSE) POLYMERASE (EC 2.4.2.30) (PARP) (ADPRT) (NAD(+)-ADP-
 CC DE RIBOSYLTRANSFERASE) (POLY(ADP-RIBOSE) SYNTHETASE).
 CC OS Sarcophaga peregrina (Flesh fly) (Boettcherisca peregrina).
 CC OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC OC Oestroidea; Sarcophagidae; Sarcophaga.
 CC NX NCBI_Taxid-7386;
 CC RN [1]
 CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 CC RX MEDLINE-94170813; PubMed-8125121;
 CC RA Masutani M., Nozaki T., Hitoi Y., Ikejima M., Nagasaki K.,
 CC de Prati A.C., Kurata S., Natori S., Sugimura T., Sami H.;
 CC RT "Cloning and functional expression of poly(ADP-ribose) polymerase
 CC RT cDNA from Sarcophaga peregrina.";
 CC RL Eur. J. Biochem. 220:607-614(1994).
 CC -1- FUNCTION: POLY(ADP-RIBOSE) POLYMERASE MODIFIES VARIOUS NUCLEAR
 CC CC PROTEINS BY POLY(ADP-RIBOSYL)ATION. THE MODIFICATION IS DEPENDENT
 CC ON DNA AND IS INVOLVED IN THE REGULATION OF VARIOUS IMPORTANT

CC CELLULAR PROCESSES SUCH AS DIFFERENTIATION, PROLIFERATION, AND
 CC TUMOR TRANSFORMATION AND ALSO IN THE REGULATION OF THE MOLECULAR
 CC EVENTS INVOLVED IN THE RECOVERY OF CELL FROM DNA DAMAGE.
 CC -1- CATALYTIC ACTIVITY: NAD(+) + (ADP-D-RIBOSYL)[N]-ACCEPTOR -
 CC NICOTINAMIDE + (ADP-D-RIBOSYL)[N+1]-ACCEPTOR.
 CC -1- COFACTOR: ZINC, CONTAINS TWO MOLES OF ZINC PER MOLE OF PROTEIN.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- MISCELLANEOUS: THE ADP-D-RIBOSYL GROUP OF NAD(+) IS TRANSFERRED TO
 CC AN ACCEPTOR CARBOXYL GROUP ON A HISTONE OR THE ENZYME ITSELF. AND
 CC THE TERMINAL ADENOSINE MOIETY, BUILDING UP A POLYMER WITH AN
 CC AVERAGE CHAIN LENGTH OF 20-30 UNITS.
 CC -1- SIMILARITY: BELONGS TO THE PARP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D16482; BAA03943.1; -
 CC HSSP: P26446; 4PAX.
 CC DR InterPro: IPR001357; BRCT.
 CC DR InterPro: IPR001290; PARP.
 CC DR InterPro: IPR001510; Znf-PARP.
 CC Pfam: PF00353; BRCT; 1.
 CC Pfam: PF00644; PARP; 1.
 CC Pfam: PF00645; Zf-PARP; 2.
 CC ProDom: PD004675; Znf-PARP; 2.
 CC SMART: SM00292; BRCT; 1.
 CC DR PROSITE: PS00172; BRCT; 1.
 CC DR PROSITE: PS00347; PARP_ZN_FINGER_1; FALSE NEG.
 CC DR PROSITE: PS00064; PARP_ZN_FINGER_2; 2.
 CC KW Transferase; Glycosyltransferase; NAD; DNA-binding; Nuclear protein;
 CC ADP-ribosylation; Zinc-finger; Zinc.
 CC FT DNA_BIND 1 369 BY SIMILARITY.
 CC FT DOMAIN 370 507 AUTOMODIFICATION DOMAIN.
 CC FT ZN_FING 19 54 NAD-BINDING.
 CC FT ZN_FING 126 164 BY SIMILARITY.
 CC FT DOMAIN 211 214 NUCLEAR LOCALIZATION SIGNAL 1ST PART.
 CC FT ACT_SITE 232 235 NUCLEAR LOCALIZATION SIGNAL 2ND PART.
 CC FT ACT_SITE 878 878 BY SIMILARITY.
 CC FT SEQUENCE 996 AA; 113018 MW; 690DD036E7487298 CRC64;
 CC -----
 CC Query Match 21.5%; Score 51.5; DB 1; Length 996;
 CC Best Local Similarity 35.1%; Pred. No. 46;
 CC Matches 13; Conservative 6; Mismatches 15; Indels 3; Gaps 1;
 CC QY 5 PCRACRLSKL---ALSICRSDFATVGRLETVLEPEE 38
 CC DB 290 PCTDCKGRQLLFHKSGYLCNGDLTEWTKYTKLKEPE 326
 CC -----
 CC RESULT 11
 CC PBP4_NOCIA STANDARD; PRT; 381 AA.
 CC ID PBP4_NOCIA
 CC AC 006317;
 CC DT 01-JUN-1994 (Rel. 29, Created)
 CC DT 01-JUN-1994 (Rel. 29, Last sequence update)
 CC DT 01-JUN-1994 (Rel. 29, Last annotation update)
 CC DE PENICILLIN-BINDING PROTEIN 4 (PBP-4).
 CC GN PBP.
 CC OS Nocardia lactudurans.
 CC OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 CC OX NCBI_Taxid-1913;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-1C411;

```

RX MEDLINE-93178438; PubMed-8440253;
RA Coque J.J.R., Liras P., Martin J.F.;
RT "Genes for a beta-lactamase, a penicillin-binding protein and a
RT transmembrane protein are clustered with the cephamycin biosynthetic
RT genes in Nocardia lactamdurans.";
RL EMBL J. 12:631-639(1993).
CC -1- FUNCTION: INVOLVED IN CELL WALL BIOSYNTHESIS AND MAY ALSO ACT
CC AS A SENSOR OF EXTERNAL PENICILLINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z13972; CAA78374.1; -.
DR PIR: S22751; S22751.
DR PIR: S22751; S22751.
KW Transmembrane; NAD; Peptidoglycan synthesis; Cell wall.
FT ACCT SITE 60 60 BY SIMILARITY.
FT TRANSMEM 271 291 POTENTIAL.
FT NP BIND 299 308 NAD (POTENTIAL).
FT TRANSDM 315 340 POTENTIAL.
SQ SEQUENCE 361 AA; 40487 MW; A3D7A80E416327EA CRC64;

Query Match 21.3%; Score 51; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 17 LSLCRSDEAIVGRLTEVL 34
: | | : | | : | |
DB 150 ISYCTGTALLIGRIYEVL 167

RESULT 12
UL47_HSV1F STANDARD; PRT; 664 AA.
AC P08113;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION PROTEIN UL47 (82/81 KDA TEGUMENT PROTEIN) (VPM82/81) (VP13/14).
GN UL47.
OS Herpes simplex virus (type 1 / strain F).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87141362; PubMed-3029433;
RA McNight J.L.C., Pellett P.E., Jenkins F.J., Roizman B.;
RT "Characterization and nucleotide sequence of two herpes simplex virus
RT 1 genes whose products modulate alpha-trans-inducing factor-dependent
RT activation of alpha genes.";
RL J. Virol. 61:992-1001(1987).
CC -1- FUNCTION: MODULATOR OF ALPHA-TIF (VPM65 PHOSPHOPROTEIN) TRANS-
CC ACTIVATION. POSSIBLY UL47 MAY HAVE KINASE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: MAJOR TEGUMENT PROTEIN OF THE VIRIONS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE LATER STAGES OF INFECTION.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL47,
CC HSV-1 13, AND VZV 11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC -----
DR EMBL: M15621; AAA45767.1; -.
DR PIR: A26133; TNBE70.
KW Transcription regulation; Trans-acting factor; Structural protein;
KW Late protein.
SQ SEQUENCE 664 AA; 70526 MW; 0EA1C56B1B7344EA CRC64;

Query Match 21.3%; Score 51; DB 1; Length 664;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 RPPPCACRLRS 13
| | | | | | | |
DB 573 RPPPCACVLRA 584

RESULT 13
CORL_HUMAN STANDARD; PRT; 1042 AA.
AC Q9Y505; Q9YHYZ;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATRIAL Natriuretic Peptide-CONVERTING ENZYME (PC 3.4.21.-) (PRO-ANP-
DE CONVERTING ENZYME) (CORIN) (HEART SPECIFIC SERINE PROTEINASE ATC2).
GN CORN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Heart;
RX MEDLINE-99262646; PubMed-10329693;
RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
RT from human heart.";
RL J. Biol. Chem. 274:14926-14935(1999).
RN [2]
RP SEQUENCE OF 734-1040 FROM N.A.
RX TISSUE-Heart;
RX MEDLINE-20534769; PubMed-11082206;
RA Hooper J.D., Scarman A.L., Clarke B.E., Normyle J.F., Antalis T.M.;
RT "Localization of the mosaic transmembrane serine protease corin to
RT heart myocytes.";
RL Eur. J. Biochem. 267:6931-6937(2000).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-20359740; PubMed-10880574;
RA Yan W., Wu F., Morser J., Wu Q.;
RT "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
RT natriuretic peptide-converting enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529(2000).
CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
CC BETWEEN ARG-123 AND SER-124.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART. EXPRESSED IN HEART
CC MYOCYTES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE
CC TRYPSTIN FAMILY.
CC -1- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FRIZZLED (FZ) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```


DR EMBL: AF13845; AAD31850.1; -
 DR EMBL: AF113248; AAF21966.1; -
 DR HSP: P01130; IAUJ.
 DR MIM: 605236; -
 DR InterPro: IPR000024; Fz_domain.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR Pfam: PF01392; Fz_2.
 DR Pfam: PF00057; Idl_recept_A; 7.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00723; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR SMART: SM00063; FRI; 2.
 DR SMART: SM00192; LDLA; 7.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS50038; Fz; 2.
 DR PROSITE: PS01209; LDLRA_1; 6.
 DR PROSITE: PS50068; LDLRA_2; 7.
 DR PROSITE: PS50240; TRYPIN_DOM; 1.
 DR PROSITE: PS00134; TRYPIN_HIS; FALSE_NEG.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; FALSE_NEG.
 DR Hydroxylase; Serine protease; Transmembrane; Signal-anchor;
 KM Glycoprotein; Repeat.
 FT DOMAIN 1 45
 FT TRANSMEM 1 66
 FT 67 1042
 FT 134 259
 FT 268 304
 FT 305 340
 FT 341 377
 FT 378 415
 FT 450 573
 FT 579 614
 FT 615 653
 FT 654 690
 FT 690 786
 FT DOMAIN 802 1042
 FT ACT_SITE 843 843
 FT ACT_SITE 892 892
 FT ACT_SITE 985 985
 FT DISULFID 790 912
 FT DISULFID 828 844
 FT DISULFID 955 970
 FT DISULFID 981 1010
 FT CARBOHYD 80 80
 FT CARBOHYD 104 104
 FT CARBOHYD 135 135
 FT CARBOHYD 141 141
 FT CARBOHYD 231 231
 FT CARBOHYD 245 245
 FT CARBOHYD 251 251
 FT CARBOHYD 305 305
 FT CARBOHYD 320 320
 FT CARBOHYD 376 376
 FT CARBOHYD 413 413
 FT CARBOHYD 446 446
 FT CARBOHYD 451 451
 FT CARBOHYD 469 469
 FT CARBOHYD 567 567
 FT CARBOHYD 651 651
 FT CARBOHYD 697 697
 FT CARBOHYD 761 761
 FT CARBOHYD 1022 1022
 FT CARBOHYD 985 985
 FT MUTAGEN 854 854
 FT CONFLICT 876 876
 SO SEQUENCE 1042 AA; 116564 MW; 7705398EBB607AD2 CRC64;

Query Match 21.3%; Score 51; DB 1; Length 1042;
 Best Local Similarity 43.2%; Pred. No. 56;
 Matches 16; Conservative 1; Mismatches 12; Indels 8; Gaps 2;

QY 2 RPPCRACRLRSKLTALSLCRDFAIVGRLEVEEPE 38
 DB 525 RPPCRALCEHRSK-----ERCESVGIYV-----LQWPE 553

RESULT 14
 ID NF1X.CHICK STANDARD; PRT; 431 AA.
 AC 090932;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NUCLEAR FACTOR 1 X-TYPE (NUCLEAR FACTOR 1/X) (NF1-X) (NF-1/X)
 DE (CCAT-BOX BINDING TRANSCRIPTION FACTOR) (CTF) (TGCA-BINDING
 DE PROTEIN).
 OS NF1X OR NF1-X.
 GN Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92093631; Pubmed=1754401;
 RA Kruse U., Qian F., Sippel A.E.;
 RT "Identification of a fourth nuclear factor I gene in chicken by cDNA
 RT cloning: NF1-X";
 RL Nucleic Acids Res. 19:6641-6641(1991).
 CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
 CC TTGCNNNNNNCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
 CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
 CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 DWA/MHI DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X61225; CAA43537.1; -
 DR InterPro: IPR000647; CTF_NF1.
 DR InterPro: IPR003619; Dwaflin_A.
 DR Pfam: PF00859; CTF_NF1; 1.
 DR SMART: SM00523; DWA; 1.
 DR PROSITE: PS00349; CTF_NF1; 1.
 KM Nuclear protein; Multigene family; Alternative splicing; Activator;
 FT NUCLEAR PROTEIN: Multigene family; Alternative splicing;
 FT DOMAIN 67 175
 FT SEQUENCE 431 AA; 46903 MW; 83989A0292E74D6C CRC64;

Query Match 21.1%; Score 50.5; DB 1; Length 431;
 Best Local Similarity 34.8%; Pred. No. 28;
 Matches 16; Conservative 7; Mismatches 16; Indels 7; Gaps 3;

QY 8 ACRLRSKLTALSL--CRSDF--AIVGRLE--VLEEPAGGIARV 46
 DB 72 ASRLAKLRDIRECREDFVLVYTGKAPCCVLSNPDQKIKIRI 117

```

RESULT 15
ID NFIC_MOUSE STANDARD; PRT; 439 AA.
AC P70255; O09072; P70256; O9RIG3;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR FACTOR 1 C-TYPE (NUCLEAR FACTOR 1/C) (NF1-C) (NF1-C) (NF1-C)
DE (CCAAT-BOX BINDING TRANSCRIPTION FACTOR) (CTF) (TGCA-BINDING
DE PROTEIN).
GN NFIC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBL_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIH SWISS;
RA Ebel T.T., Sippel A.E.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-Liver, and Skeletal muscle;
RX MEDLINE-97209336; PUBMED-9056636;
RA Chaudhry A.Z., Lyons G.E., Gronostajski R.M.;
RT "Expression patterns of the four nuclear factor I genes during mouse
RT embryogenesis indicate a potential role in development.";
RL dev. dyn. 208:313-325(1997).
RN [3]
RP SEQUENCE OF 11-187 FROM N.A.
RC STRAIN-128;
RX MEDLINE-99189145; PUBMED-10087299;
RA Fletcher C.F., Jenkins N.A., Copeland N.G., Chaudhry A.Z.,
RA Gronostajski R.M.;
RT "Exon structure of the nuclear factor I DNA-binding domain from C.
RT elegans to mammals.";
RL Mamm. Genome 10:390-396(1999).
CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
CC TGGCANNNNGCCA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
CC THE ORIGIN OF REPLICATION OF ADENOVIRUS TYPE 2. THESE PROTEINS ARE
CC INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE. LOWER
CC LEVELS IN HEART, LIVER, KIDNEY, LUNG AND BRAIN. VERY LOW LEVELS IN
CC TESTIS AND SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
CC -1- SIMILARITY: CONTAINS 1 DNA/MHI DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y07692; CAA68956.1; -
DR EMBL; U57635; AAB49930.1; -
DR EMBL; Y07693; CAA68957.1; -
DR EMBL; AF111265; AAD39100.1; -
DR MGD; MGI:109591; NF1C.
DR InterPro; IPR000647; CTF_NF1.
DR InterPro; IPR003619; Dwarfin_A.
DR Pfam; PF00859; CTF_NF1; 1.
DR SMART; SM00523; DWA; 1.
DR PROSITE; PS00349; CTF_NF1; 1.
KW Transcription regulation; DNA replication; DNA-binding; Activator;
KW Nuclear protein; Multigene family; Alternative splicing.
FT DOMAIN 68 176

```

```

FT VARSPLIC 1 10 MYSSPLICITQ -> M (IN ISOFORM 2).
SQ SEQUENCE 439 AA; 48768 MW; F794131A7A1B686A CRC64;
Query Match 21.1%; Score 50.5; DB 1; Length 439;
Best Local Similarity 34.8%; Pred. No. 29;
Matches 16; Conservative 7; Mismatches 16; Indels 7; Gaps 3;
QY 8 ACRLRSKTLASL---CRSDF--AIYGRILE--VLEPEEAGGIARY 46
DB 73 ASRLAKLRKDIRPCEDEFLVATGKAPGCVLSNPQKGMRR 118

```

Search completed: February 26, 2002, 01:46:43
Job time: 1187 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:29:39 ; Search time 78.07 Seconds
(without alignments)
45.859 Million cell updates/sec

Title: US-09-819-136-2_COPY_412_458

Perfect score: 239

Sequence: 1 PRTPPCRCACRLRSLKSLALSLC.....GRLTVELEPEEPAAGIARVA 47

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	23.8	276	2	B38965
2	56.5	23.6	423	2	E69052
3	56	23.4	3163	2	T30342
4	56	23.4	3163	2	T17440
5	54.5	22.8	446	2	F71069
6	53	22.2	147	2	S76645
7	53	22.2	188	2	T02781
8	53	22.2	266	2	H86407
9	53	22.2	363	2	E69015
10	53	22.2	378	2	C85431
11	53	22.2	941	1	A55195
12	52.5	22.0	159	2	G71289
13	52	21.8	306	2	H81036
14	52	21.8	483	2	F83205
15	52	21.8	2352	2	C83229
16	51.5	21.5	231	2	T04063
17	51.5	21.5	506	2	A33416
18	51.5	21.5	996	1	S42208
19	51.5	21.5	1141	2	A44093
20	51	21.3	291	2	T25825
21	51	21.3	381	2	S36189
22	51	21.3	664	1	TNBE70
23	50.5	21.1	220	2	E72464
24	50.5	21.1	421	2	D96641
25	50.5	21.1	431	2	S20065
26	50.5	21.1	449	2	A53562
27	50.5	21.1	475	2	T36921
28	50.5	21.1	979	2	A70848
29	50.5	21.1	1510	2	T16927

30	50	20.9	318	2	A81982	FAD synthase NMA06
31	50	20.9	471	2	T34958	probable UDP-N-ac
32	50	20.9	784	2	T45027	hypothetical prote
33	50	20.9	1634	1	JC5500	phosphoinositide 3
34	49.5	20.7	201	2	E84187	hypothetical prote
35	49.5	20.7	261	2	T51222	hypothetical prote
36	49.5	20.7	373	2	S06010	finger protein ego
37	49	20.5	132	2	F83820	cytidine/deoxycyt
38	49	20.5	151	2	T21828	hypothetical prote
39	49	20.5	182	2	T30760	hypothetical prote
40	49	20.5	214	2	C49941	devs protein - Myx
41	49	20.5	685	2	S16783	probable RNA-direc
42	49	20.5	2970	2	T08839	polyprotein - marn
43	48.5	20.3	126	1	AWHUC	natriuretic peptid
44	48.5	20.3	161	2	S55608	hypothetical prote
45	48.5	20.3	194	2	D69885	phage-related prot

ALIGNMENTS

RESULT 1

B38965 hypothetical protein B - Enterobacter agglomerans insertion sequence IS1222

C:Species: Enterobacter agglomerans

C>Date: 03-Nov-1995 #sequence-revision 03-Nov-1995 #text-change 08-Oct-1999

C:Accession: B38965; S42922

R:Steibl, H.D.; Lewecke, F.M.

Gene 156, 37-42, 1995

A>Title: IS1222: analysis and distribution of a new insertion sequence in Enterobacte

A:Reference number: A38965; MID:95255664

A:Accession: B38965

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <STE>

A:Cross-references: GB:X78052; NID:g459246; PIDN:CA54973.1; PID:g459248

A:Experimental source: strain Kleeberger, 1983

C:Genetics:

A:Mobile element: insertion sequence IS1222

C:Superfamily: Escherichia coli insertion sequence IS2 hypothetical 34k protein

Query Match 23.8% Score 57; DB 2; Length 276;

Best Local Similarity 40.5%; Pred. No. 6.2;

Matches 17; Conservative 6; Mismatches 11; Indels 8; Gaps 2;

OY 7 RACRLRSKALSLCR-----SPEAIVGRTEVLEPEAAG 41

DB 13 RACRL-TGLSLSTCRVRAORPADAHLSGRITLALERRRFG 53

RESULT 2

E69052 conserved hypothetical protein MTH1394 - Methanobacterium thermoautotrophicum (strain

C:Species: Methanobacterium thermoautotrophicum

C>Date: 03-Dec-1997 #sequence-revision 03-Dec-1997 #text-change 29-Sep-1999

C:Accession: E69052

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T

; Olu, D.; Spadafora, R.; Vicaire, J.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jilwan, K

.I.; Bacteriol. 179, 7135-7155, 1997

A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MID:98037514

A:Accession: E69052

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-423 <MTH>

A:Cross-references: GB:AE000902; GB:AE000666; NID:g2622500; PIDN:AA85871.1; PID:g2622

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1394

C:Superfamily: conserved hypothetical protein MTH1394

Query Match 23.6%; Score 56.5; DB 2; Length 423;
 Best Local Similarity 41.0%; Pred. No. 11;
 Matches 16; Conservative 5; Mismatches 13; Indels 5; Gaps 3;

OY 9 CRLRSKLAIST-CRSDFAIVGRLEVEEPP---EAAGGI 43
 DB 54 CMTREKIALSLNCRV-HEITHRIVERAMENPPRISSVGL 91

RESULT 3

T30342

protein HMP1 - *Yersinia enterocolitica*

C:Species: *Yersinia enterocolitica*

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000

C:Accession: T30342

R:pelludat, C.; Rakin, A.; Jacobi, C.; Schubert, S.; Heesemann, J.

J. Bacteriol. 180, 538-546, 1998

A:Title: The *Yersinia* adhesin biosynthetic gene cluster of *Yersinia enterocolitica*: organ

A:Reference number: 220833; MUID:96117033

A:Accession: T30342

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3161 <PEL>

A:Cross-references: EMBL:Y12527; NID:e1228641; PID:e1228642; PIDN:CAA73127.1

C:Genetics:

A:Gene: *irp1*

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:47-447/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

F:1616-1887/Domain: acyl carrier protein homology <ACPI>

F:2825-2893/Domain: acyl carrier protein homology <ACFP>

F:2856/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 23.4%; Score 56; DB 2; Length 3161;
 Best Local Similarity 46.2%; Pred. No. 84;
 Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 19 LCRSDFAIVGRLEVEEPPAAGGIA 44
 DB 1605 VCRCDVGDAGQLATVLDLDAANGGIA 1630

RESULT 4

T17440

probable polyketide synthase - *Yersinia pestis*

C:Species: *Yersinia pestis*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000

C:Accession: T17440; T47048

R:Bertherton, J.D.; Bertolino, V.T.; Perry, R.D.

Mol. Microbiol. 32, 289-299, 1999

A:Title: YbP and YbQ: two ABC transporters required for iron uptake in *Yersinia pestis*

A:Reference number: 218782; MUID:99248409

A:Accession: T17440

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3163 <PEP>

A:Cross-references: EMBL:AF091251; NID:93818595; PID:93818605; PIDN:AA069588.1

R:Buschleiser, C.; Rusniok, C.; Couve, E.; Frangeul, L.; Billault, A.; Kunst, F.; Carniel

submitted to the EMBL Data Library, October 1998

A:Description: DNA sequence of the 102 kbases unstable region of *Yersinia pestis*.

A:Reference number: 224348

A:Accession: T47048

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3163 <BUC>

A:Cross-references: EMBL:AI031866; PIDN:CAA21391.1

A:Experimental source: strain 6/69

C:Genetics:

A:Gene: *irp1*

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:47-447/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:2827-2894/Domain: acyl carrier protein homology <ACP>
 F:2858/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 23.4%; Score 56; DB 2; Length 3163;
 Best Local Similarity 46.2%; Pred. No. 84;
 Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 19 LCRSDFAIVGRLEVEEPPAAGGIA 44
 DB 1607 VCRCDVGDAGQLATVLDLDAANGGIA 1632

RESULT 5

F71069

hypothetical protein PH1251 - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: F71069

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A7100; MUID:98344137

A:Accession: F71069

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-446 <KAM>

A:Cross-references: GB:AP000005; NID:9323613; PIDN:BA03052.1; PID:93257669

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by Genba

C:Genetics:

A:Gene: PH1251

C:Superfamily: *Pyrococcus horikoshii* hypothetical protein PH1251

Query Match 22.8%; Score 54.5; DB 2; Length 446;
 Best Local Similarity 27.9%; Pred. No. 21;
 Matches 12; Conservative 7; Mismatches 13; Indels 11; Gaps 1;

OY 1 PRTPPCRACRLRSKLAISLCSDFAIVGRLEVEEPPAAGGI 43
 DB 95 PSTPLCEICGRHRHSRKPVRSEPLIG-----SGGV 126

RESULT 6

S76645

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S76645

R:Keneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamiu, E.; Nakamura, Y.; Miyajima,

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S76645

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-147 <KAN>

A:Cross-references: EMBL:D64004; GB:AB001339; NID:91001701; PIDN:BA010589.1; PID:9100

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 22.2%; Score 53; DB 2; Length 147;
 Best Local Similarity 47.6%; Pred. No. 12;
 Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 25 AIVGRLEVEEPPAAGGIAR 45
 DB 25 AIVGRLEVEEPPAAGGIAR 45

Db 35 SLTGRVADLEPPESAGGFAK 55

RESULT 7

Probable conjugal transfer protein traF precursor - Rhizobium sp. plasmid pNGR234a
C:Species: Rhizobium sp.
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C:Accession: T02781

R:Erbelberg, C.; Fellay, R.; Bairoch, A.; Broughton, W.J.; Rosenthal, A.; Perret, X.

Nature 367, 394-401, 1997

A:Title: Molecular basis of symbiosis between Rhizobium and legumes.

A:Reference number: 214734; MUID:97305956

A:Accession: T02781

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-188 <PRE>

A:Cross-References: EMBL:AE000069; NID:g2182350; PIDN:AAB91647.1; PID:g2182361

C:Genetics:

A:Gene: traF

A:Genome: plasmid pNGR234a

Query Match

Best Local Similarity 34.1%; Pred. No. 15;

Matches 14; Conservative 4; Mismatches 15; Indels 8; Gaps 1;

OY 1 PRTGPCRRLRSKSLSLC-----RSDFAIYGRLEEV 33

Db 74 PRTSAMREARMRGYLRAGLPCGCIAPLIVIAVAGORREV 114

RESULT 8

H86407

hypothetical protein AAF98436.1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: H86407

R:Rheologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chlin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marshall,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: H86407

A:Status: preliminary

A:Molecule type: DNA

R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Alredge, J.

; Liu, D.; Spadofora, R.; Vicalire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,

kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: ft

A:Reference number: A69000; MUID:98037514

A:Accession: E69015

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-363 <MTF>

A:Cross-References: GB:AE000881; GB:AE000666; NID:g2622206; PIDN:AAB85605.1; PID:g262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH116

C:superfamily: Escherichia coli glutamate decarboxylase

Query Match

Best Local Similarity 55.6%; Pred. No. 27;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 26 TVGRLEVELEPPERAGGI 43

Db 48 VIGMLGELLSEPDAACHI 65

RESULT 10

C85431

hypothetical protein AT4936530 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: C85431

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold S

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: C85431

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-378 <STO>

A:Cross-References: GB:NC_001268; NID:g7270601; PIDN:CA880319.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4936530

A:Map position: 4

Query Match

Best Local Similarity 54.2%; Pred. No. 28;

Matches 13; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

OY 5 PCRACRLRSKSLSLCSDPAIYG 28

Db 27 PARICRPRNRCIS--RRDPAIRG 48

RESULT 11

A55195

Chordin precursor - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A55195

R:Casal, Y.; Lu, B.; Steinbeisser, H.; Geisler, D.; Gont, L.K.; De Robertis, E.M.

Cell 79, 779-790, 1994

A:Title: Xenopus chordin: a novel dorsalizing factor activated by organizer-specific

A:Reference number: A55195; MUID:95094250

A:Accession: A55195

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-941 <SAS>

A:Cross-References: GB:I35764; NID:9603944; PIDN:AAC42222.1; PID:9603945

C:Superfamily: chordin; von Willebrand factor type C repeat homology <WMC1>

F:42-119/Domain: von Willebrand factor type C repeat homology <WMC2>

F:692-753/Domain: von Willebrand factor type C repeat homology <WMC2>

F:770-840/Domain: von Willebrand factor type C repeat homology <YWC3>
F:858-922/Domain: von Willebrand factor type C repeat homology <WMC4>

Query Match 22.2%; Score 53; DB 1; Length 941;
Best Local Similarity 33.5%; Pred. No. 66;
Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

OY 6 CRACRLRSKIALSLCRSDFALVGRLEVELEPEAAGIARV 45
Db 237 CGWRSILNRSTRLRMGHILVLTTLSEPEISGIKIVK 276

RESULT 12
G71289

hypothetical protein TP0707 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999

C:Accession: G71289
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770

A:Accession: G71289
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-159 <COL>

A:Cross-references: GB:AE001244; GB:AE000520; NID:93323005; PIDN:AA65694.1; PID:9332303

A:Experimental source: strain Nichols

A:Genetics:

A:Gene: TP0707
Query Match 22.0%; Score 52.5; DB 2; Length 159;
Best Local Similarity 40.0%; Pred. No. 14;
Matches 18; Conservative 5; Mismatches 19; Indels 3; Gaps 2;

OY 5 PCRCRLRSKIALSLCRSDFALVGRLEVELEPEAAGIARV 47
Db 14 PCSACTC-TQVAVQDLMSWRSGMRKRCYRRRCALMALVRVA 57

RESULT 13
H81036

riboflavin kinase/FMN adenylyltransferase NMB1834 [imported] - Neisseria meningitidis (s
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: H81036
R:Petelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizze, M.
Science 281, 1809-1815, 2000

A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappaport, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20157555

A:Accession: H81036

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-306 <TEET>

A:Cross-references: GB:AE002534; GB:AE002098; NID:97227095; PIDN:AAF42169.1; PID:9722708

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1834

C:Superfamily: conserved hypothetical protein HT0963

Query Match 21.8%; Score 52; DB 2; Length 306;
Best Local Similarity 39.6%; Pred. No. 31;
Matches 19; Conservative 4; Mismatches 21; Indels 4; Gaps 2;

OY 2 RTPPCRCRLRSKIALSLCRSDFALVGRLEVELEPEAAGIARV 46

Db 68 RMPPCRISPLRLKLELLESTGC-VDAVWLRFDPQNFSEISAGFIDRL 114

RESULT 14
F83205
probable lyase PA3516 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83205

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: F83205

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-483 <STO>

A:Cross-references: GB:AE004772; GB:AE004091; NID:99949662; PIDN:AA606904.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3516
C:Superfamily: fumarate hydratase

Query Match 21.8%; Score 52; DB 2; Length 483;
Best Local Similarity 34.8%; Pred. No. 48;
Matches 16; Conservative 7; Mismatches 11; Indels 12; Gaps 2;

OY 6 CRACRLRSKL---ALSLCRSDFALVGRLEVELEPEAAGIARV 47
Db 57 CSHCRQLRWLDVEALALAOAVGI-----PPEALETAKKA 94

RESULT 15
C83229

probable non-ribosomal peptide synthetase PA3327 [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C83229

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337

A:Accession: C83229

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2352 <STO>

A:Cross-references: GB:AE004755; GB:AE004091; NID:99949456; PIDN:AA606715.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3327

C:Superfamily: peptide synthetase ppsd; acetate--CoA ligase homology; acyl carrier pr
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:492-939/Domain: acetate--CoA ligase homology <ACLI>
F:953-1020/Domain: acyl carrier protein homology <ACPI>
F:1521-1969/Domain: acetate--CoA ligase homology <ACLI2>
F:1987-2055/Domain: acyl carrier protein homology <ACPI2>
F:985/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 21.8%; Score 52; DB 2; Length 2352;
Best Local Similarity 33.3%; Pred. No. 21e+02;
Matches 19; Conservative 5; Mismatches 17; Indels 16; Gaps 2;

OY 1 PTPPCRCRLRSKIALSLCRSDFALVGRLEVELE-----EPEAAGIA 44
Db 1206 PGPAPALACHREELA---ALASPAAVGYTWGLGAMTRLDGRCAGNEPQAAGPA 1259

Wed Feb 27 08:23:12 2002

us-09-819-136-2_copy_412_458.rpr

Page 5

Search completed: February 26, 2002, 01:29:41
Job time: 420 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:06 ; Search time 69.26 Seconds
(without alignments)
15.271 Million cell updates/sec

Title: US-09-819-136-2_COPY_412_458

Perfect score: 239
Sequence: 1 PRTPCRCRLRSKLALSLC.....GRLEVLPEEPAAGSIARVA 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.5	33.3	311	2	US-08-318-837-7
2	67.5	28.2	311	2	US-08-318-837-9
3	56	23.4	31	2	US-08-318-837-19
4	53	22.2	940	1	US-08-938-365-4
5	53	22.2	941	1	US-08-343-760A-2
6	50.5	21.1	449	2	US-08-839-008-2
7	50.5	21.1	449	2	US-08-839-008-9
8	48.5	20.3	126	3	US-07-765-830A-1
9	48.5	20.3	126	3	US-07-765-830A-6
10	48.5	20.3	126	3	US-07-765-830A-7
11	48.5	20.3	295	3	US-09-002-298-5
12	48	20.1	379	2	US-07-857-224B-87
13	47.5	19.9	20	2	US-08-318-837-4
14	47.5	19.9	377	4	US-09-420-211-2
15	47.5	19.9	861	1	US-08-484-105-18
16	47.5	19.9	861	1	US-08-484-105-18
17	47	19.7	113	4	US-09-220-528-34
18	47	19.7	116	4	US-09-220-528-35
19	47	19.7	144	4	US-09-220-528-36
20	47	19.7	185	4	US-09-220-528-41
21	47	19.7	224	4	US-09-220-528-29
22	47	19.7	970	2	US-08-673-789-7
23	47	19.7	973	1	US-08-162-809-10
24	47	19.7	988	1	US-08-162-809-14
25	46.5	19.5	126	2	US-07-757-606B-1
26	46.5	19.5	126	2	US-07-757-606B-6
27	46.5	19.5	222	1	US-08-687-895-1

28	46.5	19.5	222	2	US-09-040-482-1	Sequence 1, Appl
29	46.5	19.5	384	2	US-08-836-620A-15	Sequence 15, Appl
30	46.5	19.5	483	4	US-09-106-194-4	Sequence 4, Appl
31	46.5	19.5	483	2	US-08-836-620A-3	Sequence 3, Appl
32	46.5	19.5	548	4	US-09-139-617-1	Sequence 1, Appl
33	46.5	19.5	689	4	US-09-106-194-2	Sequence 2, Appl
34	46.5	19.5	1833	5	US-08-479-722B-2	Sequence 2, Appl
35	46.5	19.5	1833	5	PCT-US95-02251-18	Sequence 18, Appl
36	45.5	19.0	77	4	US-09-246-500B-7	Sequence 7, Appl
37	45.5	19.0	883	1	US-08-106-433A-2	Sequence 2, Appl
38	45	18.8	214	3	US-08-807-300-3	Sequence 3, Appl
39	45	18.8	295	4	US-08-893-654B-6	Sequence 6, Appl
40	45	18.8	303	4	US-09-159-106-13	Sequence 13, Appl
41	45	18.8	350	1	US-08-184-252A-2	Sequence 2, Appl
42	45	18.8	350	2	US-08-123-851-3	Sequence 3, Appl
43	45	18.8	350	2	US-08-728-520-3	Sequence 3, Appl
44	45	18.8	350	5	PCT-US95-00601-2	Sequence 2, Appl
45	45	18.8	435	4	US-09-159-106-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-318-837-7
Sequence 7, Application US/08318837
Patent No. 5981277
GENERAL INFORMATION:
APPLICANT: FRANSEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,
APPLICANT: ANDRE; VAN HEUVERSWEY, HUGO
TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID
TITLE OF INVENTION: CODING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR THERA
TITLE OF INVENTION: IMMUNOLOGY
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN AND MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,837
FILING DATE: 13-OCT-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP 93/01022
FILING DATE: 28-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 92,401,231.3
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410,007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-318-837-7
Query Match 33.3% Score 79.5; DB 2: Length 311.

Best Local Similarity 44.18; Pred. No. 0.0014;
Matches 15; Conservative 8; Mismatches 10; Indels 1; Gaps 1;
OY 5 PCRACRLSKLALSLCRSDFAIVGRLTEVLEPE 38
Db 187 PCRPCS-DTEVLALVCTSDFAVNGSIQVTHEPE 219

RESULT 2
US-08-318-837-9
; Sequence 9, Application US/08318837
; Patent No. 5981277
; GENERAL INFORMATION:
; APPLICANT: FRANKSEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,
; APPLICANT: ANDRE; VAN HEUVERSWMY, HUGO
; TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID
; TITLE OF INVENTION: COILING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR THERAPY
; TITLE OF INVENTION: IMMUNOLOGY
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318, 837
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP 93/01022
; FILING DATE: 28-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 92,401,231.3
; FILING DATE: 30-APR-1992
; TELECOMMUNICATION INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410,007
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-318-837-9

Query Match 28.28; Score 67.5; DB 2; Length 311;
Best Local Similarity 38.28; Pred. No. 0.071;
Matches 13; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

OY 5 PCRACRLSKLALSLCRSDFAIVGRLTEVLEPE 38
Db 187 PCRPCS-DTEVLALVCTSDFAVNGSIQVTHEPE 219

RESULT 3
US-08-318-837-19
; Sequence 19, Application US/08318837
; Patent No. 5981277
; GENERAL INFORMATION:
; APPLICANT: FRANKSEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,
; APPLICANT: ANDRE; VAN HEUVERSWMY, HUGO

; TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID
; TITLE OF INVENTION: COILING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR THERAPY
; TITLE OF INVENTION: IMMUNOLOGY
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318, 837
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP 93/01022
; FILING DATE: 28-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 92,401,231.3
; FILING DATE: 30-APR-1992
; TELECOMMUNICATION INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410,007
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: THP-1
US-08-318-837-19

Query Match 23.48; Score 56; DB 2; Length 31;
Best Local Similarity 52.68; Pred. No. 0.21;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 20 CRSDFAIVGRLTEVLEPE 38
Db 1 CTSDFAVNGSIQVTHEPE 19

RESULT 4
US-08-938-365-4
; Sequence 4, Application US/08938365
; Patent No. 5989909
; GENERAL INFORMATION:
; APPLICANT: Yang, Pan
; TITLE OF INVENTION: HUCHORDIN AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,365
FILING DATE: 26-SEP-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Melkilejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 09404/040001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 940 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-365-4

Query Match 22.2%; Score 53; DB 2; Length 940;
Best Local Similarity 32.5%; Pred. No. 29;
Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

QY 6 CRACRLSKLALSLCRSDFAIVGRLEVELEPEAGGIAR 45
Db 236 CGWRSLSRSTRLRLRMGHILVSLVTTLSPEISGIKIVK 275

RESULT 5
US-08-343-760A-2
Sequence 2, Application US/08343760A
Patent No. 5679783
GENERAL INFORMATION:
APPLICANT: De Robertis, Edward M
APPLICANT: Sasai, Yoshiki
TITLE OF INVENTION: Tissue Differentiation Affecting
TITLE OF INVENTION: Factor and Composition
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Majestic, Parsons, Siebert & Hsue
STREET: Four Embarcadero Center, Suite 1450
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,760A
FILING DATE: 22-NOV-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Siebert, J. Suzanne
REGISTRATION NUMBER: 28,758
REFERENCE/DOCKET NUMBER: 3100.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 363-5556
TELEFAX: (415) 362-5418
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-760A-2

Query Match 22.2%; Score 53; DB 1; Length 941;
Best Local Similarity 32.5%; Pred. No. 29;
Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

QY 6 CRACRLSKLALSLCRSDFAIVGRLEVELEPEAGGIAR 45
Db 237 CGWRSLSRSTRLRLRMGHILVSLVTTLSPEISGIKIVK 276

RESULT 6
US-08-839-008-2
Sequence 2, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baunmeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-2

Query Match 21.1%; Score 50.5; DB 2; Length 449;
Best Local Similarity 26.3%; Pred. No. 28;
Matches 10; Conservative 6; Mismatches 21; Indels 1; Gaps 1;

QY 1 PRTPPC-RACRLSKLALSLCRSDFAIVGRLEVELEPE 37
Db 313 PDAPFCPCRCRRGTIGSNFCASSILVATVAKSMVREP 350

```

RESULT      7
US-08-839-008-9
; Sequence 9, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563, 967
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-9

Query Match          21.1%; Score 50.5; DB 2; Length 449;
Best Local Similarity 26.3%; Pred. No. 28;
Matches 10; Conservative 6; Mismatches 21; Indels 1; Gaps 1.

QY      1 PRTPPC-RACRLRSKLALSLCRSDFAIVGRILEVEEHP 37
        | | | | | | | | | | : | | : :: ||
Db       313 PDAPTCPKOCRTGTGLQSNFCASSLVTATVKSMVRP 350

RESULT      8
US-07-765-830A-1
; Sequence 1, Application US/07765830A
; Patent No. 6034231
; GENERAL INFORMATION:
APPLICANT: TANAKA, SHOSI
APPLICANT: FUCHIMURA, KAYOKO
APPLICANT: TAMARAGI, YASUNORI
APPLICANT: MATSUOTO, HISAYUKI
APPLICANT: KANGAWA, KENJI
APPLICANT: MINAMINO, NAOTO
TITLE OF INVENTION: HUMAN CNP GENE AND PRECURSOR PROTEIN
NUMBER OF SEQUENCES: 8

```

01
02 CORRESPONDENCE ADDRESS:
03 ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
04 STREET: 1615 L STREET, N.W.
05 CITY: WASHINGTON
06 STATE: D.C.
07 COUNTRY: USA
08 ZIP: 20036
09
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPETER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: Patentin Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/07/765, 830A
17 FILING DATE: 19910926
18 CLASSIFICATION: 435
19
20 ATTORNEY/AGENT INFORMATION:
21 NAME: SCOTT, WATSON T
22 REGISTRATION NUMBER: 26581
23 REFERENCE/DOCKET NUMBER: 5683/92723
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 202-861-3067
26 TELEFAX: 202-822-0944
27
28 TEXEL: 6714627 CUSH
29 INFORMATION FOR SEO ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 126 amino acids
32 TYPE: AMINO ACID
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: peptide
36
37 US-07-765-830A-1

```

Query Match          20.3%; Score 48.5; DB 3, Length 126;
Best Local Similarity 31.0%; Pred. No. 12;
Matches 13; Conservative 1; Mismatches 3; Indels 25; Gaps 1;

OY      1 PRTPCRACRKSLALSLCRSDFAVGRLEVELEPEAAG 42
      |||||
      |||||
DB      32 PRTPP-----ABELAEPPQAAG 48

RESULT      9
US-07-765-830A-6
; Sequence 6, Application US/07765830A
; Patent No. 6034231
; GENERAL INFORMATION:
; APPLICANT: TANAKA, SHOSI
; APPLICANT: FUCHIMURA, KAYOKO
; APPLICANT: TAMARAGI, YASUNORI
; APPLICANT: MATSUOTO, HISAYUKI
; APPLICANT: KANGAWA, KENJI
; APPLICANT: MINAMINO, NAOTO
; TITLE OF INVENTION: HUMAN CNP GENE AND PRECURSOR PROTEIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/765,830A
; FILING DATE: 19910926
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

NAME: SCOTT, WATSON T
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 5683/92723
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3067
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-765-830A-6

Query Match 20.3%; Score 48.5; DB 3; Length 126;
Best Local Similarity 31.0%; Pred. No. 12;
Matches 13; Conservative 1; Mismatches 3; Indels 25; Gaps 1;

QY 1 PRTPPCACRLRSKLSLCSRSPFAIVGRLEVEEPAAGG 42
||||| 1 11:11111
Db 32 PRTPP-----AEELEAPQAGG 48

RESULT 10
US-07-765-830A-7
Sequence 7, Application US/07765830A
Patent No. 6034231
GENERAL INFORMATION:
APPLICANT: TANAKA, SHOSI
APPLICANT: FUCHIMURA, KAYOKO
APPLICANT: TAMARAGI, YASUNORI
APPLICANT: MATSUOTO, HISAYUKI
APPLICANT: KANGAWA, KENJI
APPLICANT: MINAMINO, NAOTO
TITLE OF INVENTION: HUMAN CNP GENE AND PRECURSOR PROTEIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/765,830A
FILING DATE: 19910926
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 5683/92723
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3067
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-765-830A-7

Query Match 20.3%; Score 48.5; DB 3; Length 126;
Best Local Similarity 31.0%; Pred. No. 12;
Matches 13; Conservative 1; Mismatches 3; Indels 25; Gaps 1;

QY 1 PRTPPCACRLRSKLSLCSRSPFAIVGRLEVEEPAAGG 42
||||| 1 11:11111
Db 32 PRTPP-----AEELEAPQAGG 48

RESULT 11
US-09-002-298-5
Sequence 5, Application US/09002298
Patent No. 6046001
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
APPLICANT: Tang, Y. Tom
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN FATY ACID BETA-OXIDATION ENZYMES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,298
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0453 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 730864
US-09-002-298-5

Query Match 20.3%; Score 48.5; DB 3; Length 295;
Best Local Similarity 43.6%; Pred. No. 33;
Matches 17; Conservative 8; Mismatches 11; Indels 3; Gaps 2;

QY 9 CRURSKLALSLCSRSPFAIVGRLEVEEPAAGIARVA 47
||||| 1 11:11111
Db 40 CRVQTE-ALVLCCKAIVGRDQERTE--QAAGISQLA 75

RESULT 12
US-07-857-224B-87
Sequence 87, Application US/07857224B
Patent No. 5958784
GENERAL INFORMATION:

DB 214 PPERIVRVIGELAVSI-GSEGLVAGQVVDVCSGMAEYGL 252

RESULT 15

US-08-484-105-18

; Sequence 18, Application US/08484105
; Patent No. 5589341

; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce

; APPLICANT: BELL, Stephen P

; APPLICANT: KOBAYASHI, Ryuj1

; APPLICANT: RINE, Jasper

; APPLICANT: FOSS, Margit

; APPLICANT: MCNALLY, Francis J

; APPLICANT: LAURENSEN, Patricia

; APPLICANT: HERSKOWITZ, Ira

; APPLICANT: LI, Joachim J

; APPLICANT: GAVIN, Kimberly

; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,105

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard Aron

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 494-8700

; TELEFAX: (415) 494-8771

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 861 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-484-105-18

Query Match 19.9%; Score 47.5; DB 1; Length 861;

Best local similarity 41.7%; Pred. No. 1.6e+02;

Matches 15; Conservative 4; Mismatches 8; Indels 9; Gaps 2;

QY 1 PTPPCRCRLSKSLALSCSDFAIVGRLEYLEE 36

DB 464 PTPPCRCRLSKSLALSCSDFAIVGRLEYLEE 490

Search completed: February 26, 2002, 01:28:07
Job time: 361 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:43 ; Search time 162.6 Seconds
(without alignments)
21.411 Million cell updates/sec

Title: US-09-819-136-2_COPY_412_458

Perfect score: 239

Sequence: 1 PRTPPCRACRLRSKLALSLC.....GRLETVLEPEAAGIARVA 47

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: A_Geneseq.1101.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.5	33.3	311	AA43682	Lipopolysaccharide
2	79.5	33.3	311	AA43683	Lipopolysaccharide
3	79.5	33.3	311	AA43684	Human acid sequenc
4	79.5	33.3	311	AA43685	Human shear stress
5	67.5	28.2	311	AA43686	Lipopolysaccharide
6	67.5	28.2	311	AA43687	Amino acid sequenc
7	57.5	24.1	22	AA43688	Amino acid sequenc
8	57	23.8	949	AA43689	Cell cycle protein
9	55.5	23.2	1247	AA43690	Cell cycle protein
10	55.5	23.2	291	AA43691	Amino acid sequenc
11	55	23.0	420	AA43692	Human protein sequ

12	54.5	22.8	716	AA43682	Amino acid sequenc
13	53	22.2	369	AA43683	Human secreted pro
14	53	22.2	369	AA43684	Human gene 11 enco
15	53	22.2	378	AA43685	Arabidopsis thalia
16	53	22.2	941	AA43686	Xenopus frog prote
17	52.5	22.0	139	AA43687	Human colon cancer
18	52	21.8	295	AA43688	Human secreted ap
19	52	21.8	295	AA43689	Amino acid sequenc
20	51.5	21.5	303	AA43690	Human apoptosis-as
21	51.5	21.5	304	AA43691	Apoptosis related
22	51.5	21.5	319	AA43692	Lung cancer associ
23	51.5	21.5	326	AA43693	Human apoptosis-as
24	51.5	21.5	366	AA43694	Human colon cancer
25	51.5	21.5	1141	AA43695	Human cerdiac CGI
26	51	21.3	1042	AA43696	Human serine prote
27	51	21.3	1042	AA43697	Human corin protei
28	50.5	21.1	334	AA43698	Gene 36 human secr
29	50.5	21.1	449	AA43699	Human smooth muscl
30	50.5	21.1	449	AA43700	Procollagen C-prot
31	50.5	21.1	449	AA43701	Human smooth muscl
32	50.5	21.1	458	AA43702	Human smooth muscl
33	49.5	20.7	40	AA43703	Human cancer assoc
34	49.5	20.7	40	AA43704	Peptide #3917 enco
35	49.5	20.7	40	AA43705	Peptide #4039 enco
36	49	20.5	148	AA43706	Peptide #3844 enco
37	49	20.5	314	AA43707	Human breast tumou
38	49	20.5	314	AA43708	Human secreted apo
39	49	20.5	328	AA43709	Human F1222-rel
40	49	20.5	328	AA43710	Human olfactory re
41	49	20.5	349	AA43711	Human OR-like poly
42	48.5	20.3	103	AA43712	Human olfactory re
43	48.5	20.3	126	AA43713	Human proCNP pepti
44	48.5	20.3	126	AA43714	hCNP. Homo saplen
45	48.5	20.3	126	AA43715	Human CNP. Homo s
			126	AA43716	Human C-type natri

ALIGNMENTS

RESULT 1
ID: AA43682 standard; Protein: 311 AA.
AC: AA43682;
DT: 24-MAY-1994 (first entry)
XX: Lipopolysaccharide induced protein.
DE: Lipopolysaccharide induced protein.
XX: Macrophage; induced; lipo-polysaccharide; antitumour;
KW: antiinflammatory; trypanocidal agent; antibody; cell proliferation;
RW: activation; cytotoxicity.
XX: Homo sapiens.
OS: Homo sapiens.
XX: W09322437-A.
XX: 11-NOV-1993.
XX: 28-APR-1993; 93WO-EP01022.
XX: 30-APR-1992; 92EP-0401231.
XX: (INNO-) INNOGENETICS NV SA.
XX: Devos K, Franssen L, Van De Voorde A, Van Heuverswyn H;
PI: WPI: 1993-368796/46.
XX: N-PSDB: AA051543.
XX: New polypeptide induced in macrophage(s) by lipo-polysaccharide -
XX: useful e.g. as antitumour, antiinflammatory or trypanocidal
XX: agent, also related nucleic acid, antibodies, anti-sense cpds.
PT

PT etc.
XX
PS Claim 9; Fig 1; 108pp; English.

CC The polypeptide induced in macrophages by lipopolysaccharide
CC stimulates cell proliferation (esp. when costimulated with IL-4)
CC promote activation, cytotoxicity, and mobilisation of LAK cells;
CC promote recruitment of suppressive peritoneal exudate cells;
CC promote generation of immunocompetent lymph node cells (LNC) and
CC have trypanocidal and trypanolytic activity. The human and murine
CC sequences are given in (AAQ51543-45), peptide fragments able to
CC generate antibodies are given in (AAR51951-61)

SQ Sequence 311 AA;

Query Match 33.3%; Score 79.5; DB 14; Length 311;
Best Local Similarity 44.1%; Pred. No. 0.0082;
Matches 15; Conservative 8; Mismatches 10; Indels 1; Gaps 1.

Dy 5 PCRACRLRSKIALSLCRSPFAIVGRLEVEEPE 38
||| |:::|||||::|::| ||
Db 187 pcrpcsdtevlavctsfavrgsqvqthpe 219

RESULT 2
AAR43683
XX AAR43683 standard; Protein; 311 AA.
XX AC AAR43683;
XX DT 24-MAY-1994 (first entry)
XX DE Lipopolysaccharide induced protein.
XX KM Macrophage; induced; lipo-polysaccharide; antitumour;
KM antiinflammatory; trypanocidal agent; antibody; cell proliferation;
KW activation; cytotoxicity.
XX OS Homo sapiens.
XX PN W09322437-A.
XX PD 11-NOV-1993.
XX PE 28-APR-1993; 93WO-EP01022.
XX PR 30-APR-1992; 92EP-0401231.
PA (INNO-) INNOGENETICS NV SA.
PI Devos K, Franssen L, Van De Voorde A, Van Heuverswyn H;
DR WPI; 1993-368796/46.
XX N-PsDB; AAQ51544.
XX PT New polypeptide induced in macrophage(s) by lipo-polysaccharide -
PT useful e.g. as antitumour; antiinflammatory or trypanocidal
PT agent, also related nucleic acid, antibodies, anti-sense cpds.
PT etc.

Claim 2; Fig 2; 108pp; English.

The polypeptide induced in macrophages by lipopolysaccharide
stimulates cell proliferation (esp. when costimulated with IL-4)
promote activation, cytotoxicity, and mobilisation of LAK cells;
promote recruitment of immunocompetent lymph node cells (LNC) and
have trypanocidal and trypanolytic activity. The human and murine
sequences are given in (AAQ51543-45), peptide fragments able to
generate antibodies are given in (AAR51951-61)

Sequence 311 AA;

QY 5 PCRACRLRSKLTALSLCNSDPAIVGRLEVELEEPE 38
 ID 187 pcrpccs-dtevlavctsfavrgsqvthepe 219
 DB 187 pcrpccs-dtevlavctsfavrgsqvthepe 219
 RESULT 3
 AAB84556
 AAB84556 standard; protein; 311 AA.
 AAB84556;
 05-SEP-2001 (first entry)
 Amino acid sequence of human SMAF-1 protein.
 Th1; Th2; Th3; cytokine; suppressive macrophage activation factor;
 SMAF-1; SMAF-2; inflammation; infection; allergy; autoimmune disease;
 transplant rejection; graft-versus host disease; malignancy;
 mucosal immunity; trypanosomiasis; inflammatory bowel disease;
 leishmaniasis; malaria; schistosomiasis; HIV-associated disease;
 measles; influenza; tuberculosis; lepra; psoriasis; multiple sclerosis;
 rheumatoid arthritis.
 Homo sapiens.
 WO200139786-A2.
 07-JUN-2001.
 20-NOV-2000; 2000WO-EP11527.
 30-NOV-1999; 99EP-0870245.
 (INNO-) INNOGENETICS NV.
 Franssen L, De Baetselier P;
 WPI: 2001-417788/44.
 New suppressive macrophage activation factor proteins, SMAF-1 or SMAF-2
 useful for the manufacture of medicament for treating type 1, type 2 or
 type 3 responses -
 Disclosure; Fig 1; 103pp; English.
 The present sequence represent human suppressive macrophage activation
 factor (SMAF)-1. SMAF-1 and SMAF-2 modulate the production of Th1, Th2
 and Th3 cytokines. The specification describes the use of SMAF-1 and
 SMAF-2 for the manufacture of a medicament for the treatment of diseases
 mediated by type 1, type 2 or type 3 responses, such as inflammation,
 infections, allergies, autoimmune diseases, transplant rejections,
 graft-versus host disease, malignancies and diseases involving mucosal
 immunity. They are used especially for the treatment of inflammatory
 bowel disease, leishmaniasis, trypanosomiasis, malaria, schistosomiasis,
 HIV-associated diseases, measles, influenza, tuberculosis, lepra, and
 infections by Candida, Borrelia, Listeria, Bordetella or Chlamydia,
 psoriasis, multiple sclerosis, and rheumatoid arthritis.
 Sequence 311 AA;

Db 187 pcrcps-dtevlavctsfavrgsqvthepe 219

RESULT 4

AAB90752 standard; Protein: 311 AA.

AC AAB90752:

DT 15-JUN-2001 (first entry)

DE Human shear stress-response protein SEQ ID NO: 4.

XX Human shear stress-response protein; vascular disease;

XX arteriosclerosis.

XX Homo sapiens.

PN WO200125427-A1.

PD 12-APR-2001.

PF 02-OCT-2000; 2000WO-JP06840.

PR 01-OCT-1999; 99JP-0280976.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI (NOJI) NOJIMA H.

PI Nojima H, Yoshitane H, Obayashi M, Ota T, Kawabata A, Sakurada K;

PI Kaga T, Sekine S, Nakamura Y, Sugano S;

DR WPI; 2001-266308/27.

DR N-PSDB; AA02875.

XX DNA sequences, proteins encoded by them and antibodies against them

PT useful in diagnosis and treatment of vascular disease caused by

PT arteriosclerosis -

XX Claim 60; Page 155-156; 678pp; Japanese.

XX The present invention provides the protein and coding sequences of a

CC number of human shear stress response proteins. These are useful in the

CC diagnosis, treatment and screening of vascular diseases caused by

CC arteriosclerosis, including heart failure, post-PTCA restenosis and

CC hypertension.

XX Sequence 311 AA;

XX Query Match 33.3%; Score 79.5; DB 22; Length 311;

XX Best Local Similarity 44.1%; Pred. NO. 0.0082;

XX Matches 15; Conservative 8; Mismatches 10; Indels 1; Gaps 1;

XX Db 5 PCRCRLRSKLTSLCRSDFAIVGRLEVEEPE 38

XX 187 pcrcps-dtevlavctsfavrgsqvthepe 219

XX RESULT 5

XX AAR43684 standard; Protein: 311 AA.

XX AC AAR43684:

XX 24-MAY-1994 (first entry)

XX Lipopolysaccharide induced protein.

XX Macrophage; induced; lipo-polysaccharide; antitumor;

XX antiinflammatory; trypanocidal agent; antibody; cell proliferation;

XX activation; cytotoxicity.

OS Mus musculus.

PN WO9322437-A.

PD 11-NOV-1993.

PF 28-APR-1993; 93WO-EP01022.

PR 30-APR-1992; 92EP-0401231.

PA (INNO-) INNOGENETICS NV SA.

PI Devos K, Franssen L, Van De Voorde A, Van Heuvelsryn H;

DR WPI; 1993-368796/46.

DR N-PSDB; AA051545.

XX New polypeptide induced in macrophage(s) by lipo-polysaccharide -

PT useful e.g. as antitumor, antiinflammatory or trypanocidal

PT agent, also related nucleic acid, antibodies, anti-sense cpds.

PT etc.

PS Claim 2; Fig 3; 108pp; English.

XX The polypeptide induced in macrophages by lipopolysaccharide

CC stimulates cell proliferation (esp. when costimulated with IL-4)

CC promote activation, cytotoxicity, and mobilisation of LAK cells;

CC promote recruitment of suppressive peritoneal exudate cells;

CC promote generation of immunocompetent lymph node cells (LNC) and

CC have trypanocidal and trypanolytic activity. The human and murine

CC sequences are given in (AA051543-45), peptide fragments able to

CC generate antibodies are given in (AAR51951-61)

XX Sequence 311 AA;

XX Query Match 28.2%; Score 67.5; DB 14; Length 311;

XX Best Local Similarity 38.2%; Pred. NO. 0.35;

XX Matches 13; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

XX Db 5 PCRCRLRSKLTSLCRSDFAIVGRLEVEEPE 38

XX 187 pcrcps-dtevlavctsfavrgsqvthepe 219

XX RESULT 6

XX AAB84555 standard; Protein: 311 AA.

XX AC AAB84555:

XX 05-SEP-2001 (first entry)

XX Amino acid sequence of murine SMAF-1 protein.

XX Th1; Th2; Th3; cytokine; suppressive macrophage activation factor;

XX SMAF-1; SMAF-2; inflammation; infection; allergy; autoimmune disease;

XX transplant rejection; graft-versus host disease; malignancy;

XX mucosal immunity; trypanosomiasis; inflammatory bowel disease;

XX leishmaniasis; malaria; schistosomiasis; HIV-associated disease;

XX measles; influenza; tuberculosis; lepra; psoriasis; multiple sclerosis;

XX rheumatoid arthritis.

XX Mus sp.

XX WO200139786-A2.

XX 07-JUN-2001.

XX 20-NOV-2000; 2000WO-EP11527.

XX 30-NOV-1999; 99EP-0870245.


```

Query Match      23.0%; Score 55; DB 22; Length 420;
Best Local Similarity 36.5%; Pred. No. 25;
Matches 19; Conservative 5; Mismatches 20; Indels 8; Gaps 2

QY 2 RTPPCRCARLRSLKSLALSLCRSD-----FAVGR-LTVEEPEAGAGIAR 45
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

[illegible]

KM tumour; neurodegenerative disorder; developmental abnormality; allergy;
 KM foetal deficiency; blood disorder; immune system disorder; arthritis;
 KM autoimmune disease; hepatic disease; renal disease; inflammation;
 KM Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
 KM infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 KM asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 KM reproductive disorder; gastrointestinal disorder; respiratory disorder;
 KM metabolic disorder; food additive; preservative; chromosome 7.
 XX
 OS Homo sapiens.
 XX
 PN WO200004140-A1.
 XX
 PD 27-JAN-2000.
 XX
 PF 14-JUL-1999; 99WO-US15849.
 XX
 PR 15-JUL-1998; 98US-0092921.
 PR 15-JUL-1998; 98US-0092922.
 PR 15-JUL-1998; 98US-0092956.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DM, Ehner R, Olsen HS, Brewer LA, Florence KA, Young PE;
 PI Mucenski M, Endress GA, Soppet DR;
 XX
 DR WPI: 2000-161128/14.
 DR N-PSDB: AA298027.
 XX
 PT New isolated human genes, useful for diagnosis and treatment of, e.g.
 PT cancers, neurological or blood disorders
 PS
 XX Disclosure: Page 448; 494pp; English.
 CC The polynucleotide sequences given in AA298017 to AA298108 encode the
 CC human secreted proteins given in AA87064 to AA87223. Human secreted
 CC protein can have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: cytostatic;
 CC immunosuppressive; antiinflammatory; neurotrophic; neuroprotective; and
 CC antiallergic. The polynucleotides and their corresponding secreted
 CC polypeptides are useful for preventing, treating or ameliorating medical
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Human secreted protein 8 and their polynucleotides can
 CC be used for developing products for the diagnosis or treatment of cancer,
 CC tumours, neurodegenerative disorders, developmental abnormalities and
 CC foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The
 CC proteins or polynucleotides can also be used as food additives or
 CC preservatives. The proteins are also useful for identifying their
 CC binding partners. AA298008 to AA298016 and AA87063 are sequence used in
 CC the exemplification of the present invention.
 XX
 SO Sequence 369 AA:
 Query Match 22.2%; Score 53; DB 21; Length 369;
 Best Local Similarity 36.4%; Pred. No. 40;
 Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
 QY 6 CRACRLSKLALSLCRSFAIV 27
 DQ 10 CQACRIKLTGPEICRMDFIVL 31
 RESULT 14

AA06146
 ID AA06146 standard; Protein: 369 AA.
 XX
 AC AA06146;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Human gene 11 encoded secreted protein fragment, SEQ ID NO:208.
 XX
 KM Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KM foetal abnormality; developmental abnormality; hematopoietic disorder;
 KM immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KM Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KM psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KM inflammation; neurological disorder; Alzheimer's disease; food additive;
 KM angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KM pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KM cell culture; chemotaxis; vunerary; binding partner identification;
 KM gene therapy; chromosome 7.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 78
 FT /note="Corresponds to any of the naturally occurring
 FT L-amino acids"
 XX
 PN WO200151504-A1.
 PD 19-JUL-2001.
 XX
 PF 12-JAN-2001; 2001WO-US00911.
 XX
 PR 13-JAN-2000; 2000US-0482273.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Komatsoulis GA, Duan DR, Rosen CA, Moore PA, Shi Y;
 PI Lafleur DM, Olsen HS, Brewer LA, Florence KA, Young PE, Soppet DR;
 PI Endress GA, Mucenski M, Ehner R;
 XX
 DR WPI: 2001-425865/45.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein 1s
 PT used in preventing, treating or ameliorating a medical condition
 PS
 XX Disclosure: Page 32; 864pp; English.
 CC AAD11630-AA011721 represent cDNAs corresponding to 71 human secreted
 CC protein genes, and AA06041-AA06132 represent the proteins they encode.
 CC AA06133-AA06205 represent human secreted protein fragments.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 71 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC hematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated

CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein fragment
CC referred to in the disclosure of the invention.

XX
SQ Sequence 369 AA;

Query Match 22.2%; Score 53; DB 22; Length 369;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

OY 6 CRACRLSKALSLCRDFAIV 27
 :::||||: ::| | |::
DB 10 cgscklktptreicrmdflvl 31

RESULT 15

AG06847
ID AAG06847 standard; Protein; 378 AA.

AC AAG06847;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 3773.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145195.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.


```
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 22.2%; Score 53; DB 21; Length 378;
Best Local Similarity 54.2%; Pred. No. 41;
Matches 13; Conservative 2; Mismatches 7; Indels 2; Gaps 1.
OY 5 PCRACRLRSKLLALSICRSDFAIVG 28
Db 27 paricrprnkceils--rrdfairg 48
```

Search completed: February 26, 2002, 01:26:45
Job time: 319 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:24 ; Search time 144.8 Seconds

(without alignments)
112.129 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_409

Perfect score: 663

Sequence: 1 CLPDVACTGCTPTSPHLVLMH.....GCGCEGNGNMFHRSCECDAC 111

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : SPTREMBL_17:*

1: SP_archaea:*

2: SP_bacteria:*

3: SP_fungi:*

4: SP_human:*

5: SP_invertebrate:*

6: SP_mammal:*

7: SP_mhc:*

8: SP_organelle:*

9: SP_phage:*

10: SP_plant:*

11: SP_prodent:*

12: SP_virus:*

13: SP_vertebrate:*

14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	249.5	37.6	3198	5 09U868	09u868 manduca sex
2	221.5	33.4	1572	5 044938	044938 haemochus
3	220.5	33.3	3060	5 09VAV4	09vav4 drosophila
4	218.5	33.0	2167	5 076840	076840 caenorhabdi
5	209	31.5	2174	5 09G0R0	09g0r0 drosophila
6	206	31.1	144	11 09OW87	09ow87 mesocricetu
7	204	30.8	151	4 P78491	P78491 homo sapien
8	194.5	29.3	287	13 093424	093424 cyprinus ca
9	194	29.3	349	11 09DB09	09db09 mus musculu
10	191	28.8	352	11 070160	070160 cavia porce
11	178.5	26.9	246	11 092208	092208 mus musculu
12	178	26.8	342	13 P70004	P70004 xenopus lae
13	175.5	26.5	2225	5 045881	045881 caenorhabdi
14	173.5	26.2	251	4 095103	095103 homo sapien
15	173.5	26.2	1965	5 061893	061893 caenorhabdi
16	171.5	25.9	922	5 021418	021418 caenorhabdi
17	170	25.6	396	6 028874	028874 canis fam11
18	165.5	25.0	1043	5 017644	017644 caenorhabdi
19	161.5	24.4	197	5 P82968	P82968 melithaea c

ALIGNMENTS

20	161	24.3	67	11	09OW86	09ow86 mesocricetu
21	158.5	23.9	1280	11	09EPX2	09epx2 mus musculu
22	158	23.8	1208	5	023456	023456 caenorhabdi
23	157.5	23.8	1743	5	09XW5	09xw5 caenorhabdi
24	157	23.7	55	5	P81902	P81902 bombyx mori
25	157	23.7	1474	5	062504	062504 caenorhabdi
26	156	23.5	58	5	09TWG0	09twg0 anemonia su
27	155.5	23.5	1235	4	095428	095428 homo sapien
28	154.5	23.3	59	5	09428	09428 anemonia su
29	154.5	23.3	507	11	09D3K4	09d3k4 mus musculu
30	154.5	23.3	507	11	099J04	099j04 mus musculu
31	154.5	23.3	1549	11	060444	060444 cricetus
32	154	23.2	58	5	09YXD3	09yxd3 fasciola he
33	152.5	23.0	838	5	027422	027422 caenorhabdi
34	152.5	23.0	838	5	018761	018761 caenorhabdi
35	151	22.8	763	5	09XZD0	09xz0 drosophila
36	150	22.6	113	5	09V508	09v508 drosophila
37	147.5	22.2	1599	5	009983	009983 caenorhabdi
38	147.5	22.2	2944	11	063870	063870 mus musculu
39	146	22.0	91	5	09VQ01	09vq01 drosophila
40	145	21.9	58	5	09TWFP9	09twf9 anemonia su
41	144	21.7	133	6	09BDL1	09bd11 macaca mula
42	144	21.7	984	5	09GQNZ	09gqnz calliacis
43	144	21.7	984	5	09GQNL	09gqnl calliacis
44	143	21.6	137	6	09BDL0	09bd10 oryctolagus
45	143	21.6	195	11	09D8Q8	09d8q8 mus musculu

RESULT 1	
09U868	PRELIMINARY; PRT: 3198 AA.
AC	09U868;
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DI	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
OS	LACUNIN PRECURSOR
OC	Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC	Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX	NCBI_Taxid=7130;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99457716; PubMed=10528409;
RA	Nardi J.B., Martos R., Walgen K.K., Lampe D.J., Robertson H.M.;
RT	"Expression of lacunin, a large multidomain extracellular matrix
RT	protein, accompanies morphogenesis of epithelial monolayers in Manduca
RT	sexta."
RL	Insect Biochem. Mol. Biol. 29:883-897(1999).
CC	-1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC	DOMAIN.
CC	-1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
CC	EMBL: AF078161; AAF04457.1; --
DR	HSSP: P12111; ZKNT.
DR	InterPro: IPR003598; IG_C2.
DR	InterPro: IPR003006; IG_MHC.
DR	InterPro: IPR002223; Kunitz_BPrt.
DR	InterPro: IPR000884; TSP1.
DR	InterPro: IPR002221; WAP.
DR	Pfam: PF00047; Ig_2.
DR	Pfam: PF00014; Kunitz_BPrt; 9.
DR	Pfam: PF00095; WAP; 1.
DR	PRINTS: PR00759; BASICPRASE.
DR	SMART: SM00408; IGc2; 2.
DR	SMART: SM00131; KU; 10.
DR	SMART: SM00209; TSP1; 7.
DR	SMART: SM00217; WAP; 1.
DR	PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
DR	PROSITE: PS00280; BPRT_KUNITZ_1; 8.
DR	PROSITE: PS00279; BPRT_KUNITZ_2; 10.


```

DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2: 3.
DR SMART: SM00131; KU: 12.
DR SMART: SM00209; TSP1: 7.
DR SMART: SM00217; WAP: 1.
DR PROSITE: PS00317; 4_DISULFIDE_CORE: 1.
DR PROSITE: PS00280; BPT1_KUNITZ_1; 11.
DR PROSITE: PS50279; BPT1_KUNITZ_2; 12.
DR PROSITE: PS50092; EGF_1; UNKNOWN_1.
DR PROSITE: PS50092; TSP1; 3.
KW Alternative splicing; Serine protease inhibitor.
FT VARSPPLIC 2803 2803 L -> SVAPV (IN SHORT ISOFORM).
FT VARSPPLIC 2844 2854 FNRKTMEDSGI -> VASPLPHNAV (IN SHORT ISOFORM).
FT VARSPPLIC 2855 3060 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3060 AA; 331579 MW; ACA31D3EE58C7C0 CRC64;

Query Match 33.3%; Score 220.5; DB 5; Length 3060;
Best Local Similarity 34.5%; Pred. No. 4.8e-15;
Matches 38; Conservative 13; Mismatches 54; Indels 5; Gaps 2;

QY 3 PDVQACTGPTSPHLVLMHYDPQGGCMFPARGCDGARGFETYEACQACARGPG-DAC 61
DB 2158 PEAGCEDNRTT---AMFYDENMACTAFYTGCGGNGNFRDRDCEGCGEFGKGVDC 2213
QY 62 VLPAVGPGRGMEPRMAYSPLLQCHPEFYVGCGEGNGNNFHSRESCEDAC 111
DB 2214 NEPVVTGPCTDMQTKYFYFMTASQACEPFYGGCGDGTGNRFSDLFECQTVAC 2263

RESULT 4
076840 PRELIMINARY; PRT; 2167 AA.
ID 076840; Q22911;
AC 076840; Q22911;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE C37C3.6 PROTEIN.
GN C37C3.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Bradshaw H.;
RT "The sequence of C. elegans cosmid C37C3."
RL Submitted (JUN-1996) to the EMBL/Genbank/DDJ databases.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC OF FORM B.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: U64857; AAC25868.1; -.
DR EMBL: U64857; AAC25867.1; -.
DR HSSP: P00981; IDTK.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR002223; Kunitz_BPT1.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00014; Kunitz_BPT1; 11.
DR Pfam: PF00090; TSP_1; 6.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2: 1.
DR SMART: SM00131; KU: 11.
DR SMART: SM00209; TSP1; 7.
DR PROSITE: PS00280; BPT1_KUNITZ_1; 10.
DR PROSITE: PS50279; BPT1_KUNITZ_2; 11.
DR PROSITE: PS50092; TSP1; 4.

```

```

KW Alternative splicing; Hypothetical protein; Serine protease inhibitor.
FT VARSPPLIC 147 155 MISSING (IN ISOFORM A).
FT VARSPPLIC 1556 1558 KRD -> SKF (IN ISOFORM A).
FT VARSPPLIC 1559 2167 MISSING (IN ISOFORM A).
SQ SEQUENCE 2167 AA; 237599 MW; 96274786D52B3639 CRC64;

Query Match 33.0%; Score 218.5; DB 5; Length 2167;
Best Local Similarity 39.8%; Pred. No. 5.7e-15;
Matches 41; Conservative 12; Mismatches 49; Indels 1; Gaps 1;

QY 9 TGPTSPHLVLMHYDPQGGCMFPARGCDGARGFETYEACQACARGPDACVLPAYG 68
DB 1453 TGPTCTNFVTKWYNKADGTCNRFHYGGCGGTNNRFPNEGCKAAC-QNHKDACQLPKYGG 1511
QY 69 PCRGMEPRMAYSPLLQCHPEFYVGCGEGNGNNFHSRESCEDAC 111
DB 1512 PCSGKHSYNYNTASHQCETFTYGGCGTGNTNRPATIEECQARC 1554

RESULT 5
09G0R0 PRELIMINARY; PRT; 2174 AA.
ID 09G0R0;
AC 09G0R0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EXTRACELLULAR MATRIX PROTEIN PAPILIN PRECURSOR.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephygroida; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephygroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DP CN BW;
RX PubMed=11076767;
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,
RA Sieron A., Prockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RT ADAMTS metalloproteinases."
RL Development 127:5475-5485(2000).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF205357; ANG37995.1; -.
DR Flybase: FBgn0003137; Ppn.
DR InterPro: IPR0003137; Ppn.
DR InterPro: IPR003599; Iq.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003600; Iq_Like.
DR InterPro: IPR002223; Kunitz_BPT1.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR002221; WAP.
DR Pfam: PF00014; Kunitz_BPT1; 3.
DR PRINTS: PR00003; 4DISULPHCORE.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2: 3.
DR SMART: SM00410; Iq_Like; 1.
DR SMART: SM00131; KU; 3.
DR SMART: SM00209; TSP1; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS50279; BPT1_KUNITZ_2; 3.
DR PROSITE: PS50092; TSP1; 3.
KW Matrix protein; Serine protease inhibitor; Signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 2174 AA; 231936 MW; 038F07952623120 CRC64;

Query Match 31.5%; Score 209; DB 5; Length 2174;
Best Local Similarity 35.7%; Pred. No. 5.9e-14;
Matches 41; Conservative 14; Mismatches 54; Indels 6; Gaps 2;

QY 3 PDVQAC-----TGPTSPHLVLMHYDPQGGCMFPARGCDGARGFETYEACQACARGP 57

```

Db 1607 POKACGLPKETGTCTNNNSVKKYFTSYGCGARFYGCGDGDNDNFSEAECKDTCQDYT 1666
 OY 58 G-DACVLPAYOGPCRGWEPKRWAYSPILQOCHFFVYGGCGGNGNNHSHSCEDAC 111
 Db 1667 GKHVLLPKRSAGPCTGFTKKWTFDYDRNRCFEGYGGCTGTTNNRPSLEQCGCTC 1721

RESULT 6

O90M87

PRELIMINARY; PRT; 144 AA.

ID 090M87
 AC 090M87
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE BIKUNIN.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN (1)
 RP SEQUENCE.
 RX MEDLINE=97018241; PubMed=8864857.
 RA Yamamoto T., Yamamoto K., Sinohara H.;
 RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian
 hamster urine and plasma."
 RL J. Biochem. 120:145-152(1996).
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR HSSP: P02760; BIK.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 KW Serine protease inhibitor
 SQ SEQUENCE 144 AA; 15700 MW; C716G1C10FC2ACES CRC64;

Query Match 31.1%; Score 206; DB 11; Length 144;
 Best Local Similarity 38.2%; Pred. No. 9.6e-15;
 Matches 39; Conservative 14; Mismatches 47; Indels 2; Gaps 1;

OY 10 GPTSPHLVLMHYDPORGCGMTFPARCGDGAARGFETYEACQACARGDACLPAVQ 69
 Db 33 GPCGLMEIKYVNGASMACFEFHYGGCLGNGNPNFSEKELQTCRTVA--ACSLPIYQGP 90
 OY 70 CRGWEPRWYSPILQOCHFFVYGGCGGNGNNHSHSCEDAC 111
 Db 91 CRAVELMAFDAAQCKVQESYGGCKGNGKFESECKEYK 132

RESULT 7

P78491

PRELIMINARY; PRT; 151 AA.

ID P78491
 AC P78491
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ALPH-1-MICROGLOBULIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91214554; PubMed=1708673;
 RA Vetr H., Gebhard W.;
 RT "Structure of the human alpha 1-microglobulin-bikunin gene."
 RL Biol. Chem. Hoppe-Seyler 371:1185-1196(1990).
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: X54818; CAA38587.1; -.

DR EMBL: X54818; CAA38587.1; JOINED.
 DR HSSP: P02760; BIK.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 KW Serine protease inhibitor.
 FT NON_TER 1
 SQ SEQUENCE 151 AA; 16542 MW; 88F400C5ECA19831 CRC64;

Query Match 30.8%; Score 204; DB 4; Length 151;
 Best Local Similarity 39.4%; Pred. No. 1.6e-14;
 Matches 41; Conservative 12; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVLMHYDPORGCGMTFPARCGDGAARGFETYEACQACARGDACLPAVQ 67
 Db 39 CMGMTSRFY---YNGTSMACETFOYGGCMGNGNNFTEKELQTCRTVA--ACNLPYR 92
 OY 68 GPCRGWEPKRWAYSPILQOCHFFVYGGCGGNGNNHSHSCEDAC 111
 Db 93 GPCRAFLQLMAFDAAVKGKCVLFPGCGQNGNKFSEKREYK 136

RESULT 8

O93424

PRELIMINARY; PRT; 287 AA.

ID O93424
 AC O93424
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 33.1 KDA PROTEIN.
 OS Cyprinus carpio (Common carp).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 CC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
 OX NCBI_TaxID=7962;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Gracey A.Y.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AF008648; AAC19410.1; -.
 DR HSSP: P31713; LSHP.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 3.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 3.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 3.
 KW Hypothetical protein; Serine protease inhibitor.
 SQ SEQUENCE 287 AA; 33093 MW; DF69B3D76718115E CRC64;

Query Match 29.3%; Score 194.5; DB 13; Length 287;
 Best Local Similarity 35.9%; Pred. No. 3.1e-13;
 Matches 37; Conservative 14; Mismatches 51; Indels 1; Gaps 1;

OY 10 GPTSPHLVLMHYDPORGCGMTFPARCGDGAARGFETYEACQAC-ARGDACLPAVQ 68
 Db 49 GPCALKADRYFPFLTDGRGCSFHYGGCGGNGNNFELQCEKCKLVKDKSPQDLDERG 108
 OY 69 PCRGWEPKRWAYSPILQOCHFFVYGGCGGNGNNHSHSCEDAC 111
 Db 109 PCRGVLPYFFDFKSOEKRFFHYGGCGGNGNNFKTIKECHERC 151

RESULT 9
 O9DBJ9 PRELIMINARY; PRT; 349 AA.

```

AC 09D839;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ALPHA 1 MICROGLOBULIN/BIRKUNIN.
GN AMBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gasterland T., Glass C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
RA Schliml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -I- Similarity: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AK04907; BAB2659.1; -.
DR MGD: MGI:88002; Ambp.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR002345; Lipocalin.
DR InterPro: IPR00566; Lipocalin_cytfabp.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR Pfam: PF00061; Lipocalin; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PRINTS: PR00179; LIPOCALIN.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KW Serine protease inhibitor.
SQ SEQUENCE 349 AA; 39101 MW; E593222FBC01BC3D CRC64;

Query Match 29.3%; Score 194; DB 11; Length 349;
Best Local Similarity 39.8%; Pred. No. 4.3e-13;
Matches 37; Conservative 13; Mismatches 41; Indels 2; Gaps 1;

QY 19 WHYDPOGCGMTFPAAGCGAARGFETYEACQACARGPGDACLPAVQPGCGMPRRA 78
DB 246 YYNAGSMACETFFQYGGCLGNGNNTFSEKDCLOTCTTA--ACNLPYVGPCHAFIKLMA 303
QY 79 YSPLLQOCHPEYVYGGCEGNGNHFHRESCEDAC 111
DB 304 FDAAGKCIQFHYGGCKGNGKFKYSEKCEKCYC 336

RESULT 10
070160 PRELIMINARY; PRT; 352 AA.
AC 070160;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ALPHA-1-MICROGLOBULIN/BIRKUNIN.
OS Cavia porcellus (Guinea pig).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hysticognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=LIVER;
RA Yoshida K., Suzuki Y., Yamamoto K., Shinohara H.;
RU Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -I- Similarity: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AB006444; BAA25305.1; -.
DR HSSP: P02760; IBIR.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR002345; Lipocalin.
DR InterPro: IPR00366; Lipocalin_cytfabp.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR Pfam: PF00061; Lipocalin; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PRINTS: PR00179; LIPOCALIN.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Serine protease inhibitor.
SQ SEQUENCE 352 AA; 39033 MW; BAA7447B574C2FAB CRC64;

Query Match 28.8%; Score 191; DB 11; Length 352;
Best Local Similarity 36.3%; Pred. No. 9e-13;
Matches 37; Conservative 13; Mismatches 50; Indels 2; Gaps 1;

QY 10 GPTSPHLVIMHYDPOGCGMTFPAAGCGAARGFETYEACQACARGPGDACLPAVQGP 69
DB 238 GPCLGLFKRFYNGSSMAEIEFYGGCLGNGNNTSEKELQTCRYVA--ACNLPYVGP 295
QY 70 CRGMEPRWAYSPLLQOCHPEYVYGGCEGNGNHFHRESCEDAC 111
DB 296 CQSSADLMAFDAIKGKCVRTTYGGCGGNGKFKYSEKCEKCYC 337

RESULT 11
094208 PRELIMINARY; PRT; 246 AA.
ID 094208;
AC 094208;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE TFP1BETA (FRAGMENT).
GN TFP1 OR TFP1BETA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Chang J.-Y., Monroe D.M., Roberts H.R.;
RU Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -I- Similarity: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF016313; AAD01586.1; -.
DR HSSP: P10646; ITFX.
DR MGD: MGI:1095418; Tfp1.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
KW Serine protease inhibitor.
FT NON_TER 1
SQ SEQUENCE 246 AA; 28003 MW; 4BBDCA23E04F07E6 CRC64;

```

Query Match 26.9%; Score 178.5; DB 11; Length 246;
 Best Local Similarity 32.0%; Pred. No. 1.4e-11;
 Matches 39; Conservative 16; Mismatches 50; Indels 17; Gaps 2;

QY 5 VQACTGPTSLVLMHYDPQGGCMTEPARGCDGAARGFETYEACQACARGP-----58
 DB 45 MKADDGCKAMINSYFLNMYTHQCEEFYIGGCEGNNRFTLDECKKTCI--FCYEKTAV 102
 QY 59 -----DACVLPAYOGCRGMEPRWASPLLOQCHPPVYGGCEGNGNNHRESCED 109
 DB 103 KAASGAEPRDCELEDDGCRGMKRYLYNQTQCCEREYVGGCLRNFFETLDECKK 162
 QY 110 AC 111
 DB 163 IC 164

RESULT 12
 P70004 PRELIMINARY; PRT; 342 AA.

AC P70004; PRELIMINARY; PRT; 342 AA.
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ALPHAI-MICROGLOBULIN/BIKUNIN PRECURSOR (AMBP).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SOURCE FROM N.A.

RC TISSUE=LIVER;
 RA Kawahara A., Hikosaka A., Sasado T., Hirota K.;
 RT "thyroid hormone-dependent repression of alpha1-microglobulin/bikunin
 RT (AMBP) gene expression during amphibian metamorphosis.";
 RL Dev. Genes Evol. 206:355-362(1997).
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: D87752; BAA13453.1; -
 DR HSSP: P02760; IBIK.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR002345; Lipocalin.
 DR InterPro: IPR00566; Lipocalin_cytfabp.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR Pfam: PF00061; Lipocalin; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR PRINTS: PR00179; LIPOCALIN.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS0279; BPTI_KUNITZ_2; 2.
 DR PROSITE: PS00213; LIPOCALIN; 1.
 KW Serine protease inhibitor; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 342 POTENTIAL.
 SO SEQUENCE 342 AA; 38558 MW; 3050508BDFC5A7B CRC64;

Query Match 26.8%; Score 178; DB 13; Length 342;
 Best Local Similarity 33.0%; Pred. No. 2.2e-11;
 Matches 34; Conservative 16; Mismatches 51; Indels 2; Gaps 1;

QY 9 TGPTSPHLVLMHYDPQGGCMTEPARGCDGAARGFETYEACQACARGPDAVLPAYOG 68
 DB 231 SGPCIGNNRKRYFNSTMACETFOYGGCLGNNNNHSEKELCHDRTET--ACRLPIPIG 288
 QY 69 PCRGMEPRWASPLLOQCHPPVYGGCEGNGNNHRESCEDAC 111
 DB 289 PCKTAKTHAFDAOGKCVTFESYGGCGNGNOFYTEKCKEYK 331

RESULT 13
 045881 PRELIMINARY; PRT; 2225 AA.

AC 045881;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE W01F3.3 PROTEIN.
 GN W01F3.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SOURCE FROM N.A.
 RA Cummings P.;
 RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SOURCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Litching J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spout J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: Z92815; CAB07294.1; -
 DR HSSP: P31713; ISHP.
 DR InterPro: IPR002899; EB.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR00716; Thyroglobulin_1.
 DR Pfam: PF00014; Kunitz_BPTI; 10.
 DR Pfam: PF00086; thyroglobulin_1; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 10.
 DR SMART: SM00211; TY; 1.
 DR SMART: SM00289; WK; 4.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 8.
 DR PROSITE: PS0279; BPTI_KUNITZ_2; 10.
 DR PROSITE: PS00484; THYROGLOBULIN_1; UNKNOWN_1.
 KW Serine protease inhibitor.
 SO SEQUENCE 2225 AA; 242197 MW; ASD8AE9D2A7B02A CRC64;

Query Match 26.5%; Score 175.5; DB 5; Length 2225;
 Best Local Similarity 31.0%; Pred. No. 2.3e-10;
 Matches 35; Conservative 9; Mismatches 56; Indels 13; Gaps 2;

QY 8 CTGPTSPHLVLMHYDPQGGCMTEPARGCDGAARGFETYEACQACARGP-----G 58
 DB 786 CRG-----QFVWPFDDKKNCDVFTYGGCGNGNNFASKECMAICHPETPTSATPDFS 841
 QY 59 DACVLPAYOGCRGMEPRWASPLLOQCHPPVYGGCEGNGNNHRESCEDAC 111
 DB 842 QVCSNDVDVAGCNGVFERFAFDAEADCRATFYGGCGGNGNNFATMDCRSRC 894

RESULT 14
 095103 PRELIMINARY; PRT; 251 AA.

AC 095103;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR BETA.
 GN TFIIBETA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RL Chang J.-Y., Monroe D.M., Roberts H.R.;
RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF021834; AAD01700.1; -
DR HSP: P10646; ITRX.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
KW Serine protease inhibitor.
SQ SEQUENCE 251 AA; 28652 MW; 9B3F276A52BAF0B9 CRC64;

Query Match 26.2%; Score 173.5; DB 4; Length 251;
Best Local Similarity 36.1%; Pred. No. 4.9e-11;
Matches 35; Conservative 11; Mismatches 38; Indels 13; Gaps 1;

OY 28 CMTFPAAGCDGAARGFETFEACQACARGPG-----DACVLPAYGPGRGWE 74
DB 79 CEEFIVGGEGNGNRFESLECKKMTNRNANRIITLQGEKPDPCFLDEPDGICRGYI 138
OY 75 PRNAYSPLAQCHPEFVYGGCEGNGNPFHSRSCEDAC 111
DB 139 TRFYNNQTKCERFKYGGCLGNMNFETLECKNIC 175

RESULT 15
ID 061893 PRELIMINARY; PRT: 1965 AA.
AC 061893;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE T07H8.4 PROTEIN.
GN T07H8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gatung S., Maggi L.;
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.

DR EMBL: AF067945; AAC17681.1; -
DR HSP: P00981; IPRK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 11.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00131; KU; 14.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 5.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 14.
DR PROSITE: PS01186; EGF_2; UNKNOWN_2.
DR PROSITE: PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat;
KW Serine protease inhibitor.
SQ SEQUENCE 1965 AA; 223504 MW; 80E22ACB1CE52234 CRC64;

Query Match 26.2%; Score 173.5; DB 5; Length 1965;
Best Local Similarity 35.8%; Pred. No. 3.4e-10;
Matches 34; Conservative 15; Mismatches 45; Indels 1; Gaps 1;

OY 18 LMHYDPRGCGMTFPAAGCDGAARGFETFEACQACARGPGDACLPAVGP-CRGMEPR 76
DB 1748 VMFYNTLRTGCDQFLYGGGNGNRFDTPEICQACEVYGTDPGMSLDGSMCEAMSNR 1807
OY 77 WAYSPLAQCHPEFVYGGCEGNGNPFHSRSCEDAC 111
DB 1808 YYENKRAQCKGFHYTGCGKSGNFLTKECCQTKC 1842

Search completed: February 26, 2002, 01:49:26
Job time: 1275 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: February 26, 2002, 01:46:43 ; Search time 45.78 Seconds

(without alignments)
88.899 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_409
Perfect score: 663
Sequence: 1 CLPDVQACTGPTSPHLVLMH.....GGCEGNGNNFHSRESCDAG 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	210	31.7	346	1	AMBP_MERUN
2	206	31.1	346	1	AMBP_MERUN
3	205	30.9	337	1	AMBP_PIG
4	205	30.9	349	1	AMBP_PIG
5	204	30.8	352	1	AMBP_HUMAN
6	200	30.2	123	1	IATR_HORSE
7	199	30.0	352	1	AMBP_BOVIN
8	197	29.7	123	1	IATR_SHEEP
9	194	29.3	349	1	AMBP_MOUSE
10	190.5	28.7	252	1	SPT2_HUMAN
11	190.5	28.7	300	1	TRP2_RABIT
12	188.5	28.4	252	1	SPT2_MOUSE
13	188	28.4	304	1	TRP1_HUMAN
14	182.5	27.5	306	1	TRP1_MOUSE
15	180.5	27.2	302	1	TRP1_RAT
16	178	26.8	304	1	TRP1_MACMU
17	174	26.2	235	1	TRP2_HUMAN
18	166	25.0	110	1	IBP_CARCR
19	165.5	25.0	62	1	IP52_ANESU
20	158.5	23.9	355	1	AMBP_PLEPL
21	158	23.8	61	1	IBP1_TACTR
22	155.5	23.5	513	1	SPT1_HUMAN
23	155	23.4	69	1	CRPT_BOOMI
24	153	23.4	1416	1	YH81_CAEL
25	150.5	22.7	507	1	SPT1_MOUSE
26	150.5	22.7	2944	1	CA17_HUMAN
27	150	22.6	134	1	TRP2_MOUSE
28	150	22.6	230	1	TRP2_MOUSE
29	149	22.5	63	1	IMAP_DROU
30	146	22.0	67	1	IBPC_BOVIN
31	146	22.0	133	1	IBP1_HUMAN
32	145	21.9	65	1	IVB3_VIPAA
33	142	21.4	58	1	AXP1_ANTAF

34	140	21.1	62	1	IVBT_ERINA
35	138	20.8	61	1	ITRS_RAT
36	138	20.8	3176	1	CA36_HUMAN
37	136	20.5	55	1	ISH2_STORE
38	136	20.5	61	1	IVB1_VIPAA
39	136	20.5	64	1	SPT3_HUMAN
40	134.5	20.3	100	1	BP22_BOVIN
41	134.5	20.3	763	1	APP2_HUMAN
42	133	20.1	53	1	HTB1_MANSE
43	133	20.1	55	1	ISH1_STORE
44	132	19.9	58	1	AXP2_ANTAF
45	131	19.8	59	1	IVBE_DENPO

ALIGNMENTS

```

RESULT 1
AMBP_MERUN
ID      AMBP_MERUN      STANDARD:      PRT:      346 AA.
AC      062577: 062576:
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE      TRYPSIN INHIBITOR LIGHT CHAIN (ITI-IC) (BIKONIN) (HI-30)].
GN      AMBP OR ITIL.
OS      Meriones unguiculatus (Mongolian jird).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Gerbillinae;
OC      Meriones.
OX      NCBI_Taxid=10047;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=95110820; PubMed=7529051;
RA      Ide H., Itoh H., Nawa Y.;
RT      "Sequencing of cDNAs encoding alpha 1-microglobulin/bikunin of
RT      Mongolian gerbil and Syrian golden hamster in comparison with man and
RT      other species."
RL      Blochin. Biophys. Acta 1209:286-292(1994).
CC      -!- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC      FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC      APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC      AND AUBUMIN (BY SIMILARITY).
CC      -!- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR. PRESENT IN PLASMA AND
CC      URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
CC      ELASTASE (BY SIMILARITY).
CC      -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC      SEPARATELY FUNCTIONING PROTEINS.
CC      -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC      FAMILY.
CC      -!- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC      -!- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb-sdb.ch).
CC      -----
CC      EMBL: D31813; BA006600.1; -.
CC      HSSP: P10646; IADZ.
DR      InterPro: IPR002223; Kunitz_BPTI.
DR      InterPro: IPR002345; Lipocalin.
DR      InterPro: IPR000566; Lipocalin_CytfAB.
DR      Pfam: PF00014; Kunitz_BPTI; 2.
DR      Pfam: PF00061; Lipocalin; 1.
DR      PRINTS: PR00179; LIPOCALIN.
DR      PRINTS: PR00759; BASICPTASE.

```

DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 Lipocalin.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
 FT CHAIN 205 346 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
 CHAIN.
 FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 FT BINDING 286 336 BPTI/KUNITZ INHIBITOR 2.
 FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 FT BINDING 90 187 BY SIMILARITY.
 FT DISULFID 230 280 BY SIMILARITY.
 FT DISULFID 239 263 BY SIMILARITY.
 FT DISULFID 255 276 BY SIMILARITY.
 FT DISULFID 286 336 BY SIMILARITY.
 FT DISULFID 295 319 BY SIMILARITY.
 FT DISULFID 311 332 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 INHIBITORY SITE (PI) (CHYMOTRYPSIN,
 ACT_SITE 240 241 ELASTASE) (BY SIMILARITY).
 FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT SEQUENCE 346 AA; 38643 MW; FLA4463810918D5F CRC64;
 SO
 Query Match 31.7%; Score 210; DB 1; Length 346;
 Best Local Similarity 39.2%; Pred. No. 3.2e-13;
 Matches 40; Conservative 13; Mismatches 47; Indels 2; Gaps 1;
 QY 10 GPISPHLVLMHYDPORGCGMTFPAKCGDGAARGEETYEACQACARPGDACYLPAYOGP 69
 DB 237 GPCIGMEKRYHYNGTSMACETFYGGCLGNGNMFISEKECLQTCRYVA--ACSLPIYOGP 294
 QY 70 CRGWEPRWVAYSPLLQCHPFGVGCCEGNGNPFHRESCEDAC 111
 DB 295 CRAVYKLMADPAQAGKCIQFYGGCKGNGKMFSEKCEKCYC 336
 RESULT 2
 AMPB_MESAU STANDARD; PRT; 349 AA.
 ID AMPB_MESAU
 AC 060559; 060558;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMPB PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
 GN AMPB OR ITIL.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95110820; PubMed=7529051;
 RA Ide H., Itoh H., Nawa Y.;
 RT "Sequencing of cDNAs encoding alpha 1-microglobulin/Bikunin of
 RT Mongolian gerbil and Syrian golden hamster in comparison with man and
 RT other species.";
 RL Biochim. Biophys. Acta 1209:286-292(1994).
 CC -!- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN (BY SIMILARITY).
 CC -!- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE (BY SIMILARITY).
 CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO

CC SEPARATELY FUNCTIONING PROTEINS.
 CC -!- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D31814; BAA06601.1; -.
 DR HSP; P10646; ITEX.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR002345; Lipocalin.
 DR InterPro; IPR000566; Lipocalin_cytfabp.
 DR Pfam; PF00014; Kunitz_BPTI; 2.
 DR Pfam; PF00061; Lipocalin_1.
 DR PRINTS; PR00179; LIPOCALIN.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 Lipocalin.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
 FT CHAIN 205 349 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
 CHAIN.
 FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 FT BINDING 286 336 BPTI/KUNITZ INHIBITOR 2.
 FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 FT DISULFID 90 187 BY SIMILARITY.
 FT DISULFID 230 280 BY SIMILARITY.
 FT DISULFID 239 263 BY SIMILARITY.
 FT DISULFID 255 276 BY SIMILARITY.
 FT DISULFID 286 336 BY SIMILARITY.
 FT DISULFID 295 319 BY SIMILARITY.
 FT DISULFID 311 332 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 INHIBITORY SITE (PI) (CHYMOTRYPSIN,
 ACT_SITE 240 241 ELASTASE) (BY SIMILARITY).
 FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPsin) (BY
 SIMILARITY).
 FT SEQUENCE 349 AA; 38782 MW; 8C954584B7DBE728 CRC64;
 SO
 Query Match 31.1%; Score 206; DB 1; Length 349;
 Best Local Similarity 38.2%; Pred. No. 7.8e-13;
 Matches 39; Conservative 14; Mismatches 47; Indels 2; Gaps 1;
 QY 10 GPISPHLVLMHYDPORGCGMTFPAKCGDGAARGEETYEACQACARPGDACYLPAYOGP 69
 DB 237 GPCIGMEKRYHYNGTSMACETFYGGCLGNGNMFISEKECLQTCRYVA--ACSLPIYOGP 294
 QY 70 CRGWEPRWVAYSPLLQCHPFGVGCCEGNGNPFHRESCEDAC 111
 DB 295 CRAVYKLMADPAQAGKCIQFYGGCKGNGKMFSEKCEKCYC 336
 RESULT 3
 AMPB_PIG STANDARD; PRT; 337 AA.
 ID AMPB_PIG
 AC P04366; P34954;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE AMBP PROTEIN PRECURSOR (CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30) (EI-14))
DE (FRAGMENT).
GN AMBP OR ITIL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
NCBI_Taxid=9823;
RN (1)
RN SEQUENCE FROM N.A.
RX MEDLINE=90353595; PubMed=1696914;
RA Gebhard W., Schreitmüller T., Veit H., Wachter E., Hochstrasser K.;
RT "Complementary DNA and deduced amino acid sequences of porcine alpha
RT 1-microglobulin and bikunin.";
RT FEBS Lett. 269:32-36(1990).
RN (2)
RN SEQUENCE OF 2-337 FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=9113729; PubMed=1703444;
RA Tavares A.;
RT "Molecular cloning of porcine alpha 1-microglobulin/HI-30 reveals
RT developmental and tissue-specific expression of two variant messenger
RT ribonucleic acids.";
RL Biochim. Biophys. Acta 1088:47-56(1991).
RN [3]
RN SEQUENCE OF 212-334.
RX MEDLINE=85323967; PubMed=2408637;
RA Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RT the inter-alpha-trypsin inhibitor, X. The amino-acid sequences of the
RT trypsin-released inhibitors from horse and pig inter-alpha-trypsin
RT inhibitors.";
RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN.
CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
CC ELASTASE.
CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: X53685; CAA37725.1; -;
DR EMBL: X52087; CAA36306.1; -;
DR PIR: A01208; TTPGRI.
DR PIR: S11066; S11066.
DR HSSP: P10646; ITFX.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000566; Lipocalin_cytfabp.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR Pfam: PF00061; Lipocalin; 1.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
DR PROSITE: PS00213; LIPOCALIN; 1.
KM Glycoprotein, Plasma, Signal, Serine protease inhibitor; Repeat;
KM Lipocalin.
FT SIGNAL 1 1
FT CHAIN <1 4 BY SIMILARITY.
FT 188 5 188 ALPHA-1-MICROGLOBULIN.

FT CHAIN 191 337 INTER-ALPHA-TRYPSIN INHIBITOR LIGHT
FT DOMAIN 216 266 CHAIN
FT DOMAIN 272 322 BPTI/KUNITZ INHIBITOR 1.
FT BINDING 38 38 BPTI/KUNITZ INHIBITOR 2.
FT DISULFID 76 173 CHROMOPHORE (BY SIMILARITY).
FT DISULFID 216 266 BY SIMILARITY.
FT DISULFID 225 249
FT DISULFID 241 262
FT DISULFID 272 322
FT DISULFID 281 305
FT DISULFID 297 318
FT CARBOHYD 100 100
FT CARBOHYD 235 235
FT ACT_SITE 226 227
FT ACT_SITE 282 283
FT CONFLICT 49 49 INHIBITORY SITE (P1) (TRYPSIN).
FT CONFLICT 259 259 T -> M (IN REF. 2).
FT CONFLICT 270 270 E -> Q (IN REF. 3).
FT CONFLICT 278 278 E -> S (IN REF. 3).
FT CONFLICT 283 283 S -> Q (IN REF. 3).
FT CONFLICT 285 286 G -> A (IN REF. 3).
FT CONFLICT 293 293 FQ -> IR (IN REF. 3).
FT CONFLICT 311 311 V -> A (IN REF. 3).
FT CONFLICT 315 315 O -> K (IN REF. 3).
SQ SEQUENCE 337 AA; 37690 MW; 1F630F983EDC70F CRC64;

Query Match 30.9%; Score 205; DB 1; Length 337;
Best Local Similarity 38.2%; Pred No. 9.5e-13;
Matches 39; Conservative 15; Mismatches 46; Indels 2; Gaps 1;

QY 10 GPTSPHLVLMYHDPOGRCCTFPARCGDGAARGFETYEACQACAPGDCVLPVAVGP 69
DB 223 GPCLGIMKIRFYNGSSMACETFYHGGCMGNGNFEVSEKELQTCR--TVACSLPIVSGP 280
QY 70 CRGWERRAYSPLLQOCHPRFYVGGCEGNGNFRSCEAC 111
DB 281 CRGFQLMAFDAVQGRCLFNYGGCGQNGMNFSEKREKVC 322

RESULT 4
ID AMBP_RAT STANDARD; PRT; 349 AA.
AC 064240; Q63336;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AMBP PROTEIN PRECURSOR (CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)).
GN AMBP OR ITIL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.
NCBI_Taxid=10116;
RN (1)
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=92182014; PubMed=1371936;
RA Lindqvist A., Bratt T., Altiieri M., Kasten W., Akerstrom B.;
RT "Rat alpha 1-microglobulin: co-expression in liver with the light
RT chain of inter-alpha-trypsin inhibitor.";
RL Biochim. Biophys. Acta 1130:63-67(1992).
RN (2)
RN SEQUENCE OF 141-195 FROM N.A.
RX MEDLINE=87033744; PubMed=2429963;
RA Kasten W., Bioerck L., Akerstrom B.;
RT "Developmental and tissue-specific expression of alpha 1-microglobulin
RT mRNA in the rat.";
RL J. Biol. Chem. 261:15070-15074(1986).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT

CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN (BY SIMILARITY).
 CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE (BY SIMILARITY).
 CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPT/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 CC EMBL: S87544; AAB21782.1; -
 CC EMBL: J02600; AAA41596.1; -
 CC HSSP: P12111; 1KUN.
 CC Interpro: IPR002223; Kunitz_BPTI.
 CC Interpro: IPR002245; Lipocalin.
 CC Interpro: IPR000566; Lipocalin_cytfabp.
 CC Pfam: PF00014; Kunitz_BPTI; 2.
 CC Pfam: PF00061; Lipocalin; 1.
 CC PRINTS: PR00179; LIPOCALIN.
 CC PRINTS: PR00759; BASICPRASE.
 CC SMART: SM00131; KU; 2.
 CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 CC PROSITE: PS00279; BPTI_KUNITZ_2; 2.
 CC PROSITE: PS00213; LIPOCALIN; 1.
 CC Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 CC Lipocalin.
 CC SIGNAL 1 19 BY SIMILARITY.
 CC CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
 CC CHAIN 205 349 INTER-ALPHA-TRYPSIN INHIBITOR LIGHT
 CC CHAIN.
 CC DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
 CC DOMAIN 286 336 BPTI/KUNITZ INHIBITOR 2.
 CC BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
 CC DISULFID 90 187 BY SIMILARITY.
 CC DISULFID 230 280 BY SIMILARITY.
 CC DISULFID 239 263 BY SIMILARITY.
 CC DISULFID 255 276 BY SIMILARITY.
 CC DISULFID 286 336 BY SIMILARITY.
 CC DISULFID 295 319 BY SIMILARITY.
 CC DISULFID 311 332 BY SIMILARITY.
 CC CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC ACT_SITE 240 241 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
 CC ELASTASE) (BY SIMILARITY).
 CC ACT_SITE 296 297 INHIBITORY SITE (P1) (TRYPSIN) (BY
 CC SIMILARITY).
 CC CONFLICT 142 142 G -> A (IN REF. 2).
 CC SEQUENCE 349 AA; 38851 MW; 1B7FB7DCB0824E01 CRC64;

Query Match 30.9%; Score 205; DB 1; Length 349;
 Best Local Similarity 39.8%; Pred. No. 9.7e-13;
 Matches 37; Conservative 14; Mismatches 40; Indels 2; Gaps 1;

QY 19 WHYDPOGCGMTFPARGDGAARCFEYEAQACARQPDGACVIVAVQGRGMEPRWA 78
 Db 246 YYNGASWACETFGYGGGLGNGNMFASKRELOTCRTIA--ACNLPYVQSCRAFAELMA 303
 QY 79 YSPLLQCHPVPVYGGCBGNGNHFHRESCEDAC 111
 Db 304 FDAAGKCIQPIYGGCKGNGKMFSEKECKEYK 336

RESULT 5
 AMBP HUMAN
 ID AMBP_HUMAN STANDARD; PRT; 352 AA.
 AC P02750; P02759; P00977;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN (PROTEIN HC)
 DE (COMPLEX-FORMING GLYCOPROTEIN HETEROGENEOUS IN CHARGE); INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-IC) (BIKUNIN) (HI-30)].
 GN AMBP OR ITIL OR HCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91214554; PubMed=1708673;
 RA Vetr H., Gebhard W.;
 RT "Structure of the human alpha 1-microglobulin-bikunin gene.";
 RL Biol. Chem. Hoppe-Seyler 371:1185-1196(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Liver;
 RX MEDLINE=87040757; PubMed=2430261;
 RA Kaumeyer J.F., Polazzi J.O., Kotick M.P.;
 RT "The mRNA for a proteinase inhibitor related to the HI-30 domain of
 RT inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin
 RT (protein HC).";
 RL Nucleic Acids Res. 14:7839-7850(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Liver;
 RX MEDLINE=90336621; PubMed=1696200;
 RA Diarra-Mentpour M., Bourguignon J., Seeboue R., Saller J.P.,
 RA Laveillard T., Martin J.P.;
 RT "Structural analysis of the human inter-alpha-trypsin inhibitor
 RT light-chain gene.";
 RL Eur. J. Biochem. 191:131-139(1990).
 RN [4]
 RP SEQUENCE OF 1-220 FROM N.A.
 RX MEDLINE=86312901; PubMed=2428011;
 RA Traboni C., Cortese R.;
 RT "Sequence of a full length cDNA coding for human protein HC (alpha 1
 RT microglobulin).";
 RL Nucleic Acids Res. 14:6340-6340(1986).
 RN [5]
 RP SEQUENCE OF 20-202 (INDIVIDUAL WITH TUBULAR PROTEINURIA).
 RX MEDLINE=84126849; PubMed=6198962;
 RA Lopez C., Grubb A.O., Mendez E.;
 RT "The complete amino acid sequence of human complex-forming
 RT glycoprotein heterogeneous in charge (protein HC) from one
 RT individual.";
 RL Arch. Biochem. Biophys. 228:544-554(1984).
 RN [6]
 RP SEQUENCE OF 20-198 (VARIANT).
 RX Lopez C., Grubb A.O., Mendez E.;
 RT "Human protein HC displays variability in its carboxyl-terminal amino
 RT acid sequence.";
 RL FEBS Lett. 144:349-353(1982).
 RN [7]
 RP SEQUENCE OF 20-198 (PATIENTS WITH TUBULAR PROTEINURIA).
 RX MEDLINE=81184038; PubMed=6164372;
 RA Takagi T., Takagi K., Kawai T.;
 RT "Complete amino acid sequence of human alpha 1-microglobulin.";
 RL Biochem. Biophys. Res. Commun. 98:997-1001(1981).
 RN [8]
 RP SEQUENCE OF 206-350.
 RX MEDLINE=85225968; PubMed=2408638;
 RA Reisinger P., Hochstrasser K., Albrecht G.J., Lempert K., Saller J.P.;
 RT "Human inter-alpha-trypsin inhibitor: localization of the Kunitz-type
 RT domains in the N-terminal part of the molecule and their release by a

RT trypsin-like proteinase.";
 RL Biol. Chem. Hoppe-Seyler 366:479-483(1985).
 RN [9]
 RP CARBOHYDRATE-LINKAGE SITES.
 RA MEDLINE-82074265; PubMed-6171497;
 RT Hochstrasser R., Schoenberger O.L., Rossmann I., Wächter E.;
 RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
 RT the inter-alpha-trypsin inhibitor. V. Attachments of carbohydrates in
 RT the human urinary trypsin inhibitor isolated by affinity
 RT chromatography.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 362:1357-1362(1981).
 RN [10]
 RP INHIBITORY SITE.
 RA MEDLINE-85225940; PubMed-3890890;
 RT Moril M., Travis J.;
 RT "The reactive site of human inter-alpha-trypsin inhibitor is in the
 RT amino-terminal half of the protein.";
 RL Biol. Chem. Hoppe-Seyler 366:19-21(1985).
 RN [11]
 RP STRUCTURE OF CARBOHYDRATES.
 RA MEDLINE-90306345; PubMed-1694784;
 RT Escaribano J., Lopez-Olin C., Hjerpe A., Mendez E.;
 RT "Location and characterization of the three carbohydrate prosthetic
 RT groups of human protein HC.";
 RL FEBS Lett. 266:167-170(1990).
 RN [12]
 RP CHROMOPHORE.
 RC TISSUE-URINE;
 RA MEDLINE-91340714; PubMed-1714898;
 RT Escaribano J., Grubb A., Calero M., Mendez E.;
 RT "The protein HC chromophore is linked to the cysteine residue at
 RT position 34 of the polypeptide chain by a reduction-resistant bond
 RT and causes the charge heterogeneity of protein HC.";
 RL J. Biol. Chem. 266:15758-15763(1991).
 RN [13]
 RP SEQUENCE OF 206-219, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
 RC TISSUE-Plasma;
 RA MEDLINE-94229087; PubMed-7513643;
 RT Morelle W., Capon C., Balduyck M., Sautiere P., Kouach M.,
 RT Michalek C., Fournet B., Mizon J.;
 RT "Chondroitin sulphate covalently cross-links the three polypeptide
 RT chains of inter-alpha-trypsin inhibitor.";
 RL Eur. J. Biochem. 221:881-888(1994).
 RN [14]
 RP SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC2.
 RA MEDLINE-93232026; PubMed-7682553;
 RT Engblid J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
 RA Pizzo S.V., Hefta S.A.;
 RT "Presence of the protein-glycosaminoglycan-protein covalent cross-link
 RT in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
 RT 2/bikunin.";
 RL J. Biol. Chem. 268:8711-8716(1993).
 RN [15]
 RP SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC3.
 RA MEDLINE-91093267; PubMed-1898736;
 RT Engblid J.J., Salvesen G., Hefta S.A., Thøgersen I.B.,
 RA Rutherford S., Pizzo S.V.;
 RT "Chondroitin 4-sulfate covalently cross-links the chains of the human
 RT blood protein pre-alpha-inhibitor.";
 RL J. Biol. Chem. 266:747-751(1991).
 RN [16]
 RP X-RAY CRYSTALOGRAPHY (2.5 ANGSTROMS) OF 230-339.
 RA MEDLINE-98227321; PubMed-9566199;
 RT Xu Y., Carr P.D., Guss J.M., Ollis D.L.;
 RT "The crystal structure of bikunin from the inter-alpha-inhibitor
 RT complex: a serine protease inhibitor with two kunitz domains.";
 RL J. Mol. Biol. 276:955-966(1998).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN.
 CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC

CC ELASTASE. ADDITIONAL PROTEOLYTIC PROCESSING IN THE KIDNEY AND/OR
 CC URINE CAN PRODUCE FURTHER AMINO- AND CARBOXYL-END MODIFICATIONS
 CC IN ITS SEQUENCE.
 CC -1- SUBUNIT: INTER-ALPHA-TRYPSIN INHIBITOR CONSIST OF A LIGHT CHAIN
 CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
 CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
 CC -1- PTM: ADDITION OF GLYCOSAMINOGLYCAN CHONDROITIN SULFATE, ALLOWS
 CC CROSS-LINKING BETWEEN THE DIFFERENT COMPONENTS.
 CC -1- MISCELLANEOUS: IN VITRO, THE FIRST TWELVE RESIDUES OF THE AMINO
 CC END OF THE INHIBITOR APPEAR TO HAVE A REACTIVE SITE CAPABLE OF
 CC INHIBITING THE ACTIVITY OF A NUMBER OF ENZYMES. ITS IN VIVO
 CC FUNCTION IS NOT KNOWN.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL: X54816; CAA38585.1; -
 CC EMBL: X54817; CAA38585.1; JOINED.
 CC EMBL: X54818; CAA38585.1; JOINED.
 CC EMBL: X04225; CAA27803.1; -
 CC EMBL: M88249; AAA59196.1; -
 CC EMBL: M88165; AAA59196.1; JOINED.
 CC EMBL: M88243; AAA59196.1; JOINED.
 CC EMBL: M88244; AAA59196.1; JOINED.
 CC EMBL: M88246; AAA59196.1; JOINED.
 CC EMBL: M88247; AAA59196.1; JOINED.
 CC EMBL: X04494; CAA28182.1; -
 CC EMBL: X54817; CAA38586.1; -
 CC PIR: A03217; HCHU
 CC PIR: A25303; A25303.
 CC PIR: S13433; S13433.
 CC PIR: S10717; S10717.
 CC PDB: 1BIK; 16-MAR-99.
 CC SWISS-2DPAGE: P02760; HUMAN.
 CC MIM: 176870; -
 CC InterPro: IPR002223; Kunitz_BPTI.
 CC InterPro: IPR002345; Lipocalin.
 CC InterPro: IPR000566; Lipocalin_cytfabp.
 CC Pfam: PF00014; Kunitz_BPTI; 2.
 CC Pfam: PF00061; Lipocalin; 1.
 CC PRINTS: PR00179; LIPOCALIN.
 CC PRINTS: PR00759; BASICPTASE.
 CC SMART: SM00131; KU; 2.
 CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 CC PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 CC PROSITE: PS00213; LIPOCALIN; 1.
 CC Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 KW
 Query Match 30.8%; Score 204; DB 1; Length 352;
 Best local similarity 39.4%; Pred. No. 1.2e-12;
 Matches 41; Conservative 12; Mismatches 45; Indels 6; Gaps 2;
 QY 8 CCGPTSLVLYWHPDQRCGCTFPARGCDGARGTEYBACQACARGDAGVPAVQ 67
 DB 240 CCMGTSRYF-----YNGTSMACEFTFGGCGMGNMNVTEKECQOTRTVA--ACNPIYR 293
 QY 68 GPGCGEPRWAVSPLDQCCHPFFYGGCEGNGNFMHSRESCDAP 111
 DB 294 GPCRAFIQLMADFAYGKCVLPFYGGCGGNGKFKYSEKRECYC 337
 RESULT 6
 IATR_HORSE

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U35642; AAB07599.1; -
 DR PIR; A01209; TTB01.
 DR HSSP; P10646; 1ADZ.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR002345; Lipocalin.
 DR InterPro: IPR000566; Lipocalin_cyFABP.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR Pfam: PF00061; Lipocalin; 1.
 DR PRINTS; PR00179; Lipocalin.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
 DR PROSITE; PS00213; Lipocalin; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 KW Lipocalin.
 FT SIGNAL 1 19
 FT CHAIN 20 203
 FT CHAIN 206 352
 FT
 FT DOMAIN 231 281
 FT BINDING 287 337
 FT DISULFID 53 53
 FT DISULFID 91 188
 FT DISULFID 231 281
 FT DISULFID 240 264
 FT DISULFID 256 277
 FT DISULFID 287 337
 FT DISULFID 296 320
 FT DISULFID 312 333
 FT ACT_SITE 241 242
 FT
 FT ACT_SITE 297 298
 FT CARBOHYD 115 115
 FT CARBOHYD 223 223
 FT CARBOHYD 250 250
 FT CONFLICT 209 209
 FT CONFLICT 217 217
 FT CONFLICT 268 268
 FT CONFLICT 274 274
 FT CONFLICT 298 299
 FT CONFLICT 330 330
 FT CONFLICT 346 346
 FT SEQUENCE 352 AA; 39233 MW; ED31C5CA02E70B19 CRC64;
 Query Match 30.0%; Score 199; DB 1; Length 352;
 Best Local Similarity 38.2%; Pred. No. 3.7e-12;
 Matches 39; Conservative 14; Mismatches 47; Indels 2; Gaps 1;
 QY 10 GPTSPHLVLMHYDPORGCMTFPAAGCDGAARGFETYEAQOACARGPDAVLPAYOGP 69
 DB 238 GPCLGKFRFYNGTSMACETFLYGGCMGNGNMFSLSEKELQICR--TVACNLPIYVGP 295
 QY 70 CRGWEPRAWYSPLOOCHPFVYGGCBGNGNMFHSRESCEDAC 111
 DB 296 CRSYIOLMARDAVKGVKCVFYSYGGCKGNGNKFYSEKCEKEYC 337
 RESULT 8
 IATR_SHEEP STANDARD; PRT; 123 AA.
 AC P13371;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE INTER-ALPHA-TRYPSIN INHIBITOR (ITI) (GIK-14) (INHIBITORY FRAGMENT OF
 DE ITI) (FRAGMENT).
 OS Ovis aries (Sheep), and Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940, 9925;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=Sheep;
 RX MEDLINE=87299012; PubMed=2441725;
 RA Rasp G., Hochstrasser K., Wachter E., Relsinger P.W.M.;
 RT "The amino-acid sequence of the trypsin-released inhibitor from sheep
 RT inter-alpha-trypsin inhibitor".
 RL Biol. Chem. Hoppe-Seyler 368:727-731(1987).
 RN [2]
 RP SEQUENCE.
 RC SPECIES=C. hircus;
 RX MEDLINE=90105540; PubMed=2481505;
 RA Rasp G., Hochstrasser K., Gerl C., Wachter E.;
 RT "Primary structure of a proteinase inhibitor released from goat serum
 RT inter-alpha-trypsin inhibitor".
 RL Biochim. Biophys. Acta 999:335-337(1989).
 CC -1- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER
 CC LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS.
 CC WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE
 CC FIRST DOMAIN INTERACTS WEAKLY WITH PAM-GRANDIOCTIC ELASTASE AND
 CC NOT AT ALL WITH PANCREATIC ELASTASE.
 CC -1- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO
 CC DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I.
 CC INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH
 CC CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 DR PIR; A29652; A29652.
 DR HSSP; P10646; 1ADZ.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam; PF00014; Kunitz_BPTI; 2.
 DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
 KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 5 55
 FT DISULFID 5 55
 FT DISULFID 14 38
 FT DISULFID 30 51
 FT DISULFID 61 111
 FT DISULFID 70 94
 FT DISULFID 86 107
 FT ACT_SITE 15 16
 FT
 FT ACT_SITE 71 72
 FT CARBOHYD 24 24
 FT NON_TER 123 123
 FT SEQUENCE 123 AA; 13686 MW; 295038173E22D2D1 CRC64;
 Query Match 29.7%; Score 197; DB 1; Length 123;
 Best Local Similarity 37.3%; Pred. No. 2.4e-12;
 Matches 38; Conservative 14; Mismatches 48; Indels 2; Gaps 1;
 QY 10 GPTSPHLVLMHYDPORGCMTFPAAGCDGAARGFETYEAQOACARGPDAVLPAYOGP 69
 DB 12 GPCLGKFRFYNGTSMACETFLYGGCMGNGNMFPSKELQICR--TVACNLPIYVGP 69
 QY 70 CRGWEPRAWYSPLOOCHPFVYGGCBGNGNMFHSRESCEDAC 111
 DB 70 CRAGIELMARDAVKGVKCVFYSYGGCKGNGNQFYSEKCEKEYC 111
 RESULT 9
 AMBP_MOUSE

ID AMBP_MOUSE STANDARD; PRT: 349 AA.
AC 007456; Q61294;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
GN AMBP OR ITIL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/C; TISSUE=Liver;
RX MEDLINE=93363639; PubMed=7689339;
RA Chan P., Sailer J.P.;
RT "Mouse alpha-1-microglobulin/Bikunin precursor: cDNA analysis, gene
RT evolution and physical assignment of the gene next to the orosomucoid
RT locus";
RL Biochim. Biophys. Acta 1174:195-200(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=95189774; PubMed=7533761;
RA Itoh H., Ide H., Katoaka H., Tomita M., Yoshinara H., Nawa Y.;
RT "cDNA sequencing of mouse alpha 1-microglobulin/inter-alpha-trypsin
RT inhibitor light chain and its expression in acute inflammation";
RL J. Biochem. 116:767-772(1994).
RN [3]
RP SEQUENCE OF 128-349 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RA Itoh H., Ide H., Yoshinara H., Nawa Y.;
RT Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN (BY SIMILARITY).
CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
CC ELASTASE (BY SIMILARITY).
CC -1- SUBUNIT: INTER-ALPHA-TRYPSIN INHIBITOR CONSIST OF A LIGHT CHAIN
CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
CC SIMILARITY).
CC -1- SIMILARITY: IN THE N-TERMINAL LINKED BROWN-YELLOW CHROMOPHORE (BY
CC FAMILY).
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X68680; CAA48640.1; -
DR EMBL: D28812; BAA05973.1; -
DR HSSP: P12111; IKUN.
DR MGD: MGI:88002; Ambp.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR002345; Lipocalin.
DR InterPro: IPR000566; Lipocalin_cytfABP.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR Pfam: PF00061; Lipocalin; 1.
DR PRINTS: PR00179; LIPOCALIN.
DR PRINTS: PR00759; BASICPASE.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 2.

DR PROSITE: PS00213; LIPOCALIN; 1.
KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
KW Lipocalin.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
FT CHAIN 205 349 INTER-ALPHA-TRYPSIN INHIBITOR LIGHT
FT CHAIN
FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
FT DOMAIN 286 336 BPTI/KUNITZ INHIBITOR 2.
FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
FT DISULFID 90 187 BY SIMILARITY.
FT DISULFID 230 280 BY SIMILARITY.
FT DISULFID 239 263 BY SIMILARITY.
FT DISULFID 255 276 BY SIMILARITY.
FT DISULFID 266 336 BY SIMILARITY.
FT DISULFID 295 319 BY SIMILARITY.
FT DISULFID 311 332 BY SIMILARITY.
FT CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 INHIBITORY SITE (PI) (CHYMOTRYPSIN,
FT ACT_SITE 240 241 ELASTASE) (BY SIMILARITY).
FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPSIN) (BY
FT ACT_SITE
FT CONFLICT 65 65 Q -> S (IN REF. 2).
FT SEQUENCE 349 AA; 39070 MM; CEAD9FC7375DAB0B CRC64;
SQ
Query Match 29.3%; Score 194; DB 1; Length 349;
Best local Similarity 39.8%; Pred. No. 1,1e-11;
Matches 37; Conservatve 13; Mismatches 41; Indels 2; Gaps 1;
QY 19 WHYDPQRGCGMTTPARCGDCAAGFEETEECCQACARCPDCAVLAPOPCGSGEP RNA 78
Db 246 YYYGASMACETFEYGGCLGNNGNFFISEKDLQTCRTIA-ACNLPTVDSCHRAFTK LMA 303
QY 79 YSPLLQCHPEYVGGCGEGNGNPFRESCEAD 111
Db 304 FDAAGKCIOPHYGGCKGNGKRFSECKEYEC 336
RESULT 10
SPT2_HUMAN STANDARD; PRT: 252 AA.
AC 043291; 000271; O14895;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR
DE ACTIVATOR INHIBITOR TYPE 2) (HAI-2) (PLACENTAL BIKUNIN).
GN SPTN2 OR HAI2 OR KOP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96010584; PubMed=9346890;
RA Kawaguchi T., Qin L., Shimomura T., Kondo J., Matsumoto K., Denda K.,
RA Kitamura N.;
RT "Purification and cloning of hepatocyte growth factor activator
RT inhibitor type 2, a Kunitz-type serine protease inhibitor";
RL J. Biol. Chem. 272:27558-27564(1997).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-74.
RC TISSUE=Placenta;
RX MEDLINE=97273732; PubMed=9115294;
RA Marlor C.W., Delaria K.A., Davis G., Muller D.K., Greve J.M.,
RA Tamburini P.P.;
RT "Identification and cloning of human placental bikunin, a novel serine
RT protease inhibitor containing two Kunitz domains";
RL J. Biol. Chem. 272:12202-12208(1997).
RN [3]

RP SEQUENCE FROM N.A.
 RC TISSUE-pancreatic cancer;
 RX MEDLINE-98094245; PubMed-9434156;
 RA Mueller-Pillasch F., Wallrapp C., Bartels K., Varga G., Friess H.,
 RA Buechler M., Adler G., Gress T.M.;
 RT "Cloning of a new kunitz-type protease inhibitor with a putative
 RT transmembrane domain overexpressed in pancreatic cancer.";
 RL Blochim. Biophys. Acta 1395:88-95(1998).
 CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR. ALSO INHIBITS PLASMIN,
 CC PLASMA AND TISSUE KALLIKREIN, AND FACTOR XIA.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, KIDNEY, PANCREAS,
 CC PROSTATE, TESTIS, THYMUS, AND TRACHEA.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB06534; BAA25024.1; -
 CC EMBL: U78095; AAC02781.1; -
 CC EMBL: AF027205; AAB84031.1; -
 CC HSSP: P05067; ITAM.
 DR MIM: 605124; -
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI_2.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 KW Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;
 KW Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.
 FT DOMAIN 28 197 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 198 218 POTENTIAL.
 FT DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
 FT DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
 FT DISULFID 38 88 BY SIMILARITY.
 FT DISULFID 47 71 BY SIMILARITY.
 FT DISULFID 63 84 BY SIMILARITY.
 FT ACT_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 133 183 BY SIMILARITY.
 FT DISULFID 142 166 BY SIMILARITY.
 FT DISULFID 158 179 BY SIMILARITY.
 FT ACT_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 54 57 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 3 3 Q -> H (IN REF. 3).
 FT CONFLICT 11 11 R -> P (IN REF. 1).
 FT CONFLICT 53 53 R -> K (IN REF. 3).
 FT CONFLICT 240 240 D -> H (IN REF. 3).
 SQ SEQUENCE 252 AA; 28228 MW; A7D3360CDECBAB2B CRC64;

Query Match 28.7%; Score 190.5; DB 1; Length 252;
 Best Local Similarity 29.1%; Pred. No. 1.8e-11;
 Matches 43; Conservative 12; Mismatches 54; Indels 39; Gaps 3;

OY 1 CLPDAVQACTPTSPHLVLMHYDPQRCGCMTPFARGCDGAARGFETFEACQACA----- 54
 DB 38 CL-VSKVVGRCRASMPRMWYNTVTDSCQLFYVGGCDGNSNLTREBECLKAKATYENA 95
 OY 55 -----KGPDA-----CVLPAVQGPCRGMEPRMAYSPLL 83
 DB 96 TCDLATSRMAADSVSPADRRDSEDSHSDMFNEYECYANAVTGCRCASFPWMYFDVER 155

OY 84 OOCHEPVYGCCEGNGNHNHRESCEDAC 111
 DB 156 NSCNMTIYGCRCNKRNSYSEACMLRC 183
 RESULT 11
 ID TFEI_RABIT STANDARD; PRT; 300 AA.
 AC P19761; Q28828; 01-FEB-1991 (rel. 17, Created)
 DT 01-AUG-1992 (rel. 23, Last sequence update)
 DE 20-AUG-2001 (rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)
 DE (EPI).
 GN TFEI.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_Taxid=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Liver;
 RX MEDLINE-91057146; PubMed-2136251;
 RA Wesselschmidt R.L., Girard T.J., Broze G.J. Jr.;
 RT "cDNA sequence of rabbit lipoprotein-associated coagulation
 RT inhibitor."; Nucleic Acids Res. 18:6440-6440(1990).
 RN [2]
 RP REVISIONS TO 72; 211 AND 218.
 RC TISSUE-Liver;
 RX MEDLINE-92335027; PubMed-1630940;
 RA Warn-Cramer B.-J., Broze G.J. Jr., Komlives E.A.;
 RT "cDNA sequence of rabbit tissue factor pathway inhibitor."; Nucleic Acids Res. 20:3548-3548(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Lung;
 RX MEDLINE-93276427; PubMed-8503123;
 RA Belaaouaj A., Kuppuswamy M.N., Birktoft J.J., Bajaj S.P.;
 RT "Revised cDNA sequence of rabbit tissue factor pathway inhibitor."; Thromb. Res. 69:547-553(1993).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
 CC A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN
 CC ANTITHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 CC LIPOPROTEINS IN PLASMA.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X54708; CAA38515.1; ALT_SEQ.
 CC EMBL: S61902; AAB26836.1; -
 DR PIR: S12143; S12143.
 DR HSSP: P10646; ITFX.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI_3.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 3.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 3.
 KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 KW Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 300 TISSUE FACTOR PATHWAY INHIBITOR.

FT DOMAIN 50 100 BPTI/KUNITZ INHIBITOR 1
 FT DOMAIN 121 171 (VII(A)/TISSUE FACTOR BINDING SITE).
 FT DOMAIN 213 263 BPTI/KUNITZ INHIBITOR 2
 FT (BY SIMILARITY).
 FT DISULFID 50 100 BY SIMILARITY.
 FT DISULFID 59 83 BY SIMILARITY.
 FT DISULFID 75 96 BY SIMILARITY.
 FT ACT_SITE 60 61 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 121 171 BY SIMILARITY.
 FT DISULFID 130 154 BY SIMILARITY.
 FT ACT_SITE 146 167 REACTIVE BOND (BY SIMILARITY).
 FT ACT_SITE 131 132 BY SIMILARITY.
 FT DISULFID 213 263 BY SIMILARITY.
 FT DISULFID 222 246 BY SIMILARITY.
 FT DISULFID 238 259 REACTIVE BOND (BY SIMILARITY).
 FT ACT_SITE 223 224 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 31 31 MISSING (IN REF. 3).
 FT CONFLICT 269 272 PKSI -> RNLS (IN REF. 3).
 SQ SEQUENCE 300 AA; 34435 MM; A08DE36537708CA6 CRC64;

Query Match 28.7%; Score 190.5; DB 1; Length 300;
 Best Local Similarity 34.8%; Pred. No. 2.1e-11;
 Matches 40; Conservative 15; Mismatches 47; Indels 13; Gaps 1;

OY 10 GPTSPHLVLMWHPDQRCGCMTEFPARGCDGARGFEYEAQQACAR-----G 56
 DB 57 GPCRAVTRFFNFNITTHGCEFTYGGCGNENRFESECEKAKADPKMTKLPFGK 116
 OY 57 PGDACVLPVAVGPGCRGWEPRMAYSPLLQQCPHYVGGCGNGNNFHSRESCDAG 111
 DB 117 KPDEFLEEDPGICRGYITRFYNNOSKOCERFKYGGGLGNINNESLECKNTC 171

RESULT 12
 SPT2_MOUSE STANDARD; PRT: 252 AA.

ID SPT2_MOUSE
 AC Q9WU03; Q9WU04; Q9WU05;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR
 DE ACTIVATOR INHIBITOR TYPE 2) (HA1-2).
 GN SPINT2 OR HA12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN-BALB/C;
 RA MEDLINE=99160423; PubMed=10049781;
 RA Itoh H., Kataoka H., Hamasuna R., Kitamura N., Koono M.;
 RA "Hepatocyte growth factor activator inhibitor type 2 lacking the first
 RA Kunitz-type serine proteinase inhibitor domain is a predominant
 RA product in mouse but not in human."
 RL Biochem. Biophys. Res. Commun. 255:740-748(1999).
 CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC -1- PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: ISOFORM 2 IS MORE PREDOMINANTLY EXPRESSED THAN
 CC ISOFORM 1.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AF099016; AAD22172.1; -
 CC EMBL: AF099019; AAD22173.1; -
 CC EMBL: AF099020; AAD22174.1; -
 CC HSSP: P05067; ITAM.
 CC MGD: MGI:138051; Spint2.
 CC InterPro: IPR002223; Kunitz-BPTI.
 CC Pfam: PF00014; Kunitz-BPTI; 2.
 CC PRINTS: PR00759; BASICPTASE.
 CC SMART: SM00131; KU; 2.
 CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 CC PROSITE: PS00279; BPTI_KUNITZ_2; 2.
 CC Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;
 CC Signal; Alternative splicing.
 CC SIGNAL 1 27 POTENTIAL.
 CC CHAIN 28 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.
 CC DOMAIN 28 197 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 198 218 POTENTIAL.
 CC DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
 CC DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
 CC DISULFID 38 88 BY SIMILARITY.
 CC DISULFID 47 71 BY SIMILARITY.
 CC DISULFID 63 84 BY SIMILARITY.
 CC ACT_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
 CC DISULFID 133 183 BY SIMILARITY.
 CC DISULFID 142 166 BY SIMILARITY.
 CC DISULFID 158 179 REACTIVE BOND (BY SIMILARITY).
 CC ACT_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
 CC CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 37 93 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 CC VARSPLIC 114 128 PKRSADLSAEITFN -> CFVLSVAALFIPYA (IN
 CC VARSPLIC 129 252 MISSING (IN ISOFORM 3).
 CC VARSPLIC 252 AA; 27914 MM; B2F4B86924D4F8F CRC64;

Query Match 28.4%; Score 188.5; DB 1; Length 252;
 Best Local Similarity 28.5%; Pred. No. 2.8e-11;
 Matches 41; Conservative 12; Mismatches 54; Indels 37; Gaps 2;

OY 5 VQACGPTSPHLVLMWHPDQRCGCMTEFPARGCDGARGFEYEAQQACAR----- 54
 DB 40 VSKYVGGKCRASIPRMWYNTTDSQCPFYGGCEBNGNNYOSKECDLKAGVTENTIDM 99
 OY 55 --RGPGDA-----CVLPVAVGPGCRGWEPRMAYSPLLQQCH 87
 DB 100 ARNRGAGSSVLSVPRKQSAEDLSAEIPNYEYECVPAVATGPCRAAPRMWYDTEKNSCI 159
 OY 88 PFYVGGCGNGNNFHSRESCDAG 111
 DB 160 SEIYGGCRGNKNSYLSQECMQHC 183

RESULT 13
 TFPI_HUMAN STANDARD; PRT: 304 AA.

ID TFPI_HUMAN
 AC P10646;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)
 DE (EPI).
 GN TFPI OR TFPI1 OR LACI.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88198127; PubMed=2452157;
 RA Mun T.-C., Kretzmer K.K., Girard T.J., Miletich J.P., Broze G.J. Jr.;
 RT "Cloning and characterization of a cDNA coding for the lipoprotein-
 associated coagulation inhibitor shows that it consists of three
 tandem Kunitz-type inhibitory domains.";
 RL J. Biol. Chem. 263:6001-6004(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129227; PubMed=1993173;
 RA van der Logt C.P.E., Reitsma P.H., Bertina R.M.;
 RT "Intron-exon organization of the human gene coding for the
 lipoprotein-associated coagulation inhibitor: the factor Xa dependent
 inhibitor of the extrinsic pathway of coagulation.";
 RL Biochemistry 30:1571-1577(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91161593; PubMed=2002045;
 RA Girard T.J., Eddy R., Messelschmidt R.L., Macphail L.A.,
 RA Likert K.M., Byers M.G., Shows T.B., Broze G.J. Jr.;
 RT "Structure of the human lipoprotein-associated coagulation inhibitor
 gene. Intron/exon gene organization and localization of the gene to
 chromosome 2.";
 RL J. Biol. Chem. 266:5036-5041(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89388722; PubMed=2781520;
 RA Girard T.J., Warren L.A., Novotny W.F., Bejcek B.E., Miletich J.P.,
 Broze G.J. Jr.;
 RT "Identification of the 1.4 kb and 4.0 kb messages for the lipoprotein
 associated coagulation inhibitor and expression of the encoded
 protein.";
 RL Thromb. Res. 55:37-50(1989).
 RN [5]
 RP SEQUENCE OF 28-50.
 RX MEDLINE=90036996; PubMed=2553722;
 RA Novotny W.F., Girard T.J., Miletich J.P., Broze G.J. Jr.;
 RT "Purification and characterization of the lipoprotein-associated
 coagulation inhibitor from human plasma.";
 RL J. Biol. Chem. 264:18832-18837(1989).
 RN [6]
 RP INHIBITORY SITES.
 RX MEDLINE=89181950; PubMed=2927510;
 RA Girard T.J., Warren L.A., Novotny W.F., Likert K.M., Brown S.G.,
 RA Miletich J.P., Broze G.J. Jr.;
 RT "Functional significance of the Kunitz-type inhibitory domains of
 lipoprotein-associated coagulation inhibitor.";
 RL Nature 338:518-520(1989).
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=96224851; PubMed=8639592;
 RA Nakahara Y., Miyata T., Hamuro T., Funatsu A., Miyagi M.,
 Tsunawasa S., Kato H.;
 RT "Amino acid sequence and carbohydrate structure of a recombinant
 human tissue factor pathway inhibitor expressed in Chinese hamster
 ovary cells: one N- and two O-linked carbohydrate chains are located
 between Kunitz domains 2 and 3 and one N-linked carbohydrate chain is
 in Kunitz domain 2.";
 RL Biochemistry 35:6450-6459(1996).
 RN [8]
 RP REVIEW.
 RX MEDLINE=91104709; PubMed=2271516;
 RA Broze G.J. Jr., Girard T.J., Novotny W.F.;
 RT "Regulation of coagulation by a multivalent Kunitz-type inhibitor.";
 RL Biochemistry 29:7539-7546(1990).
 RN [9]
 RP STRUCTURE BY NMR OF 121-182.
 RX MEDLINE=97342711; PubMed=9199408;
 RA Burgerling M.J., Orbons L.P., van der Doelen A., Mulders J.,
 Theunissen H.J., Grootehuis P.D., Bode W., Huber R., Stubbs M.T.;

RT "The second Kunitz domain of human tissue factor pathway inhibitor:
 cloning, structure determination and interaction with factor Xa.";
 RL J. Mol. Biol. 269:395-407(1997).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
 A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN
 ANTITHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 LIPOPROTEINS IN PLASMA.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: MOSTLY IN ENDOTHELIAL CELLS.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- PTM: O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib.ch).
 CC -----
 CC EMBL J03225; AAA52022.1; -
 CC EMBL M58650; AAA59480.1; -
 CC EMBL M58644; AAA59480.1; JOINED.
 CC EMBL M58645; AAA59480.1; JOINED.
 CC EMBL M58646; AAA59480.1; JOINED.
 CC EMBL M58647; AAA59480.1; JOINED.
 CC EMBL M58648; AAA59480.1; JOINED.
 CC EMBL M58649; AAA59480.1; JOINED.
 CC EMBL M59489; AAA59526.1; -
 CC EMBL M59493; AAA59526.1; JOINED.
 CC EMBL M59494; AAA59526.1; JOINED.
 CC EMBL M59495; AAA59526.1; JOINED.
 CC EMBL M59496; AAA59526.1; JOINED.
 CC EMBL M59497; AAA59526.1; JOINED.
 CC EMBL M59498; AAA59526.1; JOINED.
 CC PIR; A28650; TTHOCK.
 CC PIR; A34315; A34315.
 CC PIR; A60433; A60433.
 CC PIR; S03903; S03903.
 CC PDB; 1ADZ; 25-FEB-98.
 CC PDB; 1TFX; 21-JAN-98.
 CC GlycositeDB; P10646; -.
 CC MIM; 152310; -.
 CC InterPro; IPR002223; Kunitz_BPTI.
 CC Pfam; PF00014; Kunitz_BPTI; 3.
 CC PRINTS; PR00759; BASICPTASE.
 CC SMART; SM00131; KU; 3.
 CC PROSITE; PS00280; BPTI_KUNITZ_1; 3.
 CC PROSITE; PS0279; BPTI_KUNITZ_2; 3.
 CC Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 CC Signal; 3D-structure.
 CC SIGNAL 1 28
 CC CHAIN 29 304
 CC DOMAIN 54 104
 CC DOMAIN 125 175
 CC DOMAIN 217 267
 CC DISULFID 54 104
 CC DISULFID 63 87
 CC DISULFID 79 100
 CC ACT_SITE 64 65
 CC DISULFID 125 175
 CC DISULFID 134 158
 CC DISULFID 150 171
 CC ACT_SITE 135 136
 CC DISULFID 217 267
 CC DISULFID 226 250
 CC DISULFID 242 263
 CC ACT_SITE 227 228
 CC CARBOHD 145 145
 CC -----
 CC TISSUE FACTOR PATHWAY INHIBITOR.
 CC BPTI/KUNITZ INHIBITOR 1
 CC (VII(A)/TISSUE FACTOR BINDING SITE).
 CC BPTI/KUNITZ INHIBITOR 2
 CC (FACTOR X(A) BINDING SITE).
 CC BPTI/KUNITZ INHIBITOR 3.
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC REACTIVE BOND (BY SIMILARITY).
 CC BY SIMILARITY.
 CC REACTIVE BOND (BY SIMILARITY).
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC REACTIVE BOND (BY SIMILARITY).
 CC BY SIMILARITY.
 CC REACTIVE BOND (BY SIMILARITY).
 CC BY SIMILARITY.
 CC N-LINKED (GLCNAC. . .).

FT CARBOHYD 202 202 O-LINKED.
 FT CARBOHYD 203 203 O-LINKED.
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .).
 FT MURGEN 64 64 R->I: ABOLISHES INHIBITION OF VII(A)/TF.
 FT MURGEN 135 135 R->L: ABOLISHES INHIBITION OF X(A).
 FT MURGEN 227 227 R->L: ABOLISHES INHIBITION OF VII(A)/TF.
 SQ SEQUENCE 304 AA; 35015 MW; 5281E32B758B44FE CRC64;

Query Match 28.4%; Score 188; DB 1; Length 304;
 Best Local Similarity 29.6%; Pred. No. 3.6e-11;
 Matches 42; Conservative 14; Mismatches 46; Indels 38; Gaps 2;

OY 4 DVQACTGPTSPHLVLMHYDPORGCGMTTPARCGDGAARGFETYEACQACARPG-----58
 DB 130 DPGICRG-----YTRFYVYNQTKOCERFKYGGCLGNMNFETLECKNICEGPNQPOVD 185
 OY 59 -----DACVLPANVGPCRGWEPNAYSPILQOCHPF 89
 DB 186 NYGTOLANVNSLTPQSTKRVPSLFEFHGPSWCLTPADRGGLCRANENRFRYNSVIGICKRPF 245
 OY 90 VYGCCEGNGNMFHRESCDACC 111
 DB 246 KYSGCGENNTSKQECRLRAC 267

RESULT 14
 TFPI_MOUSE STANDARD; PRT; 306 AA.

AC 054819;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR
 DE (EPI)).
 GN TFPI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=98152575; PubMed=9493581;
 RA Chang J.-Y., Monroe D.M., Oliver J.A., Liles D.K., Roberts H.R.;
 RT "Cloning, expression, and characterization of mouse tissue factor
 RT pathway inhibitor (TFPI).";
 RL Thromb. Haemost. 79:306-309(1998).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
 CC A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN
 CC ANTIHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 CC LIPOPROTEINS IN PLASMA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF004833; AAC40035.1; -
 CC HSSP: P10646; 1TFX.
 CC MGD: MGI:1095418; Tfp1.
 CC InterPro: IPR002223; Kunitz_BPTI.
 CC Pfam: PF00014; Kunitz_BPTI; 3.
 CC PRINTS: PR00759; BASICPTASE.
 CC SMART: SM00131; KU; 3.

DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 3.
 KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 FT SIGNAL.
 FT CHAIN 1 28 BY SIMILARITY.
 FT DOMAIN 29 306 TISSUE FACTOR PATHWAY INHIBITOR.
 FT DOMAIN 50 100 BPTI/KUNITZ INHIBITOR 1
 FT DOMAIN 121 171 (VII(A)/TISSUE FACTOR BINDING SITE).
 FT DOMAIN 121 171 BPTI/KUNITZ INHIBITOR 2
 FT DOMAIN 121 171 (FACTOR X(A) BINDING SITE).
 FT DOMAIN 121 171 BPTI/KUNITZ INHIBITOR 3.
 FT DISULFID 50 100 BY SIMILARITY.
 FT DISULFID 59 83 BY SIMILARITY.
 FT DISULFID 75 96 BY SIMILARITY.
 FT ACT_SITE 60 61 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 121 171 BY SIMILARITY.
 FT DISULFID 130 154 BY SIMILARITY.
 FT DISULFID 146 167 REACTIVE BOND (BY SIMILARITY).
 FT ACT_SITE 131 132 BY SIMILARITY.
 FT DISULFID 225 275 BY SIMILARITY.
 FT DISULFID 234 258 BY SIMILARITY.
 FT DISULFID 250 271 BY SIMILARITY.
 FT ACT_SITE 235 236 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 306 AA; 34987 MW; D3EA3297E4B6A359 CRC64;

Query Match 27.5%; Score 182.5; DB 1; Length 306;
 Best Local Similarity 32.0%; Pred. No. 1.2e-10;
 Matches 39; Conservative 17; Mismatches 49; Indels 17; Gaps 2;

OY 5 VQACTGPTSPHLVLMHYDPORGCGMTTPARCGDGAARGFETYEACQACARPG-----58
 DB 52 MKADDGCKAMIRSFNMTTHQCEEFYGGCGENRFPDLECKKTCI--DGYERTAV 109
 OY 59 -----DACVLPANVGPCRGWEPNAYSPILQOCHPRVYGGCEGNGNPHRESCD 109
 DB 110 KASGARPDPFCTLEDPDGLCRGIMKRYLYNOTKOCERFVGGCLGNMNFETLECKR 169
 OY 110 AC 111
 DB 170 IC 171

RESULT 15
 TFPI_RAT STANDARD; PRT; 302 AA.

AC 002445;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR
 DE (EPI)).
 GN TFPI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=92348361; PubMed=1639767;
 RA Enjujoi K.-I., Emt M., Mukai T., Kato H.;
 RT "cDNA cloning and expression of rat tissue factor pathway inhibitor
 RT (TFPI).";
 RL J. Biochem. 111:681-687(1992).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
 CC A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN

[illegible]

Search completed: February 26, 2002, 01:46:43
Job time: 1187 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:29:41 ; Search time 78.07 Seconds
(without alignments)
108.305 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_409
Perfect score: 663
Sequence: 1 CLPDDVQACTPTSPHLVLMH.....GCCEGNGNHFHRSCECDAC 111

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	218.5	33.0	2167	T34395	hypothetical prote
2	205	30.9	337	T1PGBI	alpha-1-microglobu
3	205	30.9	349	S21089	alpha-1-microglobu
4	204	30.8	352	1 HCHU	alpha-1-microglobu
5	200	30.2	125	1 TTHOBI	alpha-1-microglobu
6	199	30.0	352	1 TTHOBI	alpha-1-microglobu
7	197	29.7	123	A29652	inter-alpha-trypsi
8	194	29.3	349	S35708	inter-alpha-trypsi
9	190.5	28.7	300	S12143	tissue factor path
10	190.5	28.7	300	S12143	lipoprotein-assoc
11	188.5	28.4	252	1 TGO185	hepatocyte growth
12	188	28.4	304	1 TTHUGK	tissue factor path
13	180.5	27.2	302	1 TTRTKG	tissue factor path
14	178	26.8	304	1 JC2264	tissue factor path
15	175.5	26.5	2225	2 T26063	hypothetical prote
16	174	26.2	235	A54951	tissue factor path
17	173.5	26.2	1965	2 T33216	hypothetical prote
18	171.5	25.9	922	2 T25373	hypothetical prote
19	170	25.6	396	2 S53325	tissue factor path
20	166	25.0	110	1 TITTOR	basic proteinase 1
21	165.5	25.0	62	S07451	hypothetical prote
22	165.5	25.0	1043	2 T19734	hypothetical prote
23	158.5	23.9	355	1 S22181	gamma-1-microglobu
24	158	23.8	61	1 TTRCBP	proteinase inhibit
25	158	23.8	1208	2 T21822	hypothetical prote
26	158	23.8	2844	2 T26859	hypothetical prote
27	157.5	23.8	1743	2 T26859	hypothetical prote
28	154.5	23.3	1549	2 T48103	type VII collagen
29	152.5	23.0	838	2 T20125	hypothetical prote

30	150.5	22.7	2944	2 A54849	collagen alpha 1(V
31	149	22.5	63	1 T1PFAF	acrosin inhibitor
32	147.5	22.2	1599	2 T16210	hypothetical prote
33	146	22.0	67	1 T1BOC	trypsin inhibitor,
34	145	21.9	57	2 A59204	basic proteinase 1
35	145	21.9	62	2 A44180	taicatoxin serine
36	145	21.9	65	1 T1VIVC	venom basic protei
37	141.5	21.3	1391	2 T20406	hypothetical prote
38	141.5	21.3	2150	2 T32497	hypothetical prote
39	140	21.1	62	2 S19327	venom basic protei
40	139	21.0	372	2 JC2556	alpha-1-microglobu
41	138	20.8	3176	1 CGHU3A	collagen alpha 3(V
42	136	20.5	61	1 T1VITL	venom basic protei
43	136	20.5	64	2 S41399	kunitz-type protei
44	134.5	20.3	100	1 T1BOSP	spleen basic prote
45	134.5	20.3	751	2 A49974	beta-amyloid precu

ALIGNMENTS

```

RESULT 1
T34395
hypothetical protein C37C3.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
C:Accession: T34395; T34394
R:Geisels, C.; Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid C37C3.
A:Reference number: 221518
A:Accession: T34395
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b
A:Experimental source: strain Bristol N2; clone C37C3
A:Accession: T34394
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1555, 'SKF' <GE2>
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a
A:Experimental source: strain Bristol N2; clone C37C3
C:Genetics:
A:Gene: CESP:C37C3.6b; CESP:C37C3.6a
A:Map position: 5
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 15
Query Match 33.0%; Score 218.5; DB 2; Length 2167;
Best Local Similarity 39.8%; Pred. No. 7.3e-13;
Matches 41; Conservative 12; Mismatches 49; Indels 1; Gaps 1;
OY 9 TGPTSPHLVLMHYDPGRGCGMPFARGCGAARGFETVACQACGARGGDCVLPAYOG 68
DB 1453 TGECTFVTKMYTKADGTCNRFHGGCGGTNNRFPNEQCKKAC-QNHKRCQLPKYVG 1511
OY 69 PCRGWEPRAWSPLLQOCHPEYVGGCEGNGNHFHRSCECDAC 111
DB 1512 PCGKSHSYVYVWASHQCEFTYGGCTLGTMNRPATIECOARC 1554
RESULT 2
T1PGBI
alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor - pig (fragment)
N:Alternate names: bikunin; IRI; PI-14 (inhibitory fragment of IRI)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 30-Jun-1987 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: S11066; S13493; A01208
R:Gebhard, W.; Schreitmuller, T.; Vetr, H.; Wachter, E.; Hochstrasser, K.
FEBS Lett. 269, 32-36, 1990
A:Title: Complementary DNA and deduced amino acid sequences of porcine alpha1-microg
A:Reference number: S11066; MUID:90353595

```

A:Accession: S11066
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-337 <GEN>
 A:Cross-references: EMBL:X53685; NID:q1877; PIDN:CAA37725.1; PID:q1878
 R:Yavakol, A.
 Biochim. Biophys. Acta 1088, 47-56, 1991
 A:Title: Molecular cloning of porcine alpha(1)-microglobulin/HI-30 reveals developmental
 A:Reference number: S13493; MUID:91113729
 A:Accession: S13493
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'W',3-48,'M',50-337 <TAV>
 A:Cross-references: GB:X52087; NID:q1881; PIDN:CAA36306.1; PID:q1882
 A:Note: The authors translated the codon GTG for residue 2 as a Met initiation codon
 R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Reisinger, P.
 Biol. Chem. Hoppe-Seyler 366, 473-478, 1985
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-
 A:Reference number: A50685; MUID:85225967
 A:Accession: A01208
 A:Molecule type: protein
 A:Residues: 212-258,'Q',260-269,'S',271-277,'Q',279-282,'A',284,'IR',287-292,'A',294-310
 C:Comment: This inhibitory fragment, released from native ITI after limited proteolysis
 first domain interacts weakly with PMN-granulocytic elastase and not at all with pancrea
 nd elastase; those with leucine interact strongly.
 C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocalin h
 C:Keywords: duplication; glycoprotein; plasma; serine proteinase inhibitor
 F:20-173/Domain: lipocalin homology <LIP>
 F:216-266/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:272-332/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 F:216-266,225-249,241-262,272-322,281-305,297-318/Disulfide bonds: #status predicted
 F:226/Inhibitory site: Leu (chymotrypsin, elastase) #status predicted
 F:233/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:282/Inhibitory site: Arg (trypsin) #status predicted

Query Match 30.9%; Score 205; DB 1; Length 337;
 Best Local Similarity 36.2%; Pred. No. 2,6e-12;
 Matches 39; Conservative 15; Mismatches 46; Indels 2; Gaps 1;

OY 10 GPTSPHLYLMHYDPORGCMTFPARGCGARGFETYEACQACARCGDVCYLPAYOGR 69
 DB 223 GPGCGMTRKYRYNNGSSNACETFFHGGCGNGNNGNFFSEKECLQTR--TVKACSLPISGPR 280
 OY 70 CRGWEPRWVSPLLQGCHEFYVGGCEGNGNMFHSRSCEDAC 111
 DB 281 CRGFOLMAEDAVOQKCVLEFNGGCGNGNGNFYSEKECKEYC 322

RESULT 3
 S21089
 alpha-1-microglobulin/inter-alpha-trypsin inhibitor light chain precursor - rat
 N:Alternate names: acid-stable proteinase inhibitor; bikunin; trypstatin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 22-Nov-1993 #sequence,revision 01-Sep-1995 #text,change 04-Feb-2000
 C:Accession: S21089; A53056; A25935; A31890; A61633
 R:Lindqvist, A.; Bretz, T.; Altieri, M.; Kastern, W.; Akerstrom, B.
 Biochim. Biophys. Acta 1130, 63-67, 1992
 A:Title: Rat alpha(1)-microglobulin: co-expression in liver with the light chain of inte
 A:Reference number: S21089; MUID:92182014
 A:Accession: S21089
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-349 <LIN>
 A:Cross-references: GB:S87544; NID:q247162; PIDN:AAB21782.1; PID:q247163
 R:Itoh, H.; Ide, H.; Ishikawa, N.; Nawa, Y.
 J. Biol. Chem. 269, 3818-3822, 1994
 A:Title: Mast cell protease inhibitor, trypstatin, is a fragment of inter-alpha-trypsin
 A:Reference number: A53056; MUID:94148892
 A:Accession: A53056
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA

A:Residues: 285-341 <ITO>
 R:Kastern, W.; Bjorck, L.; Akerstrom, B.
 J. Biol. Chem. 261, 15070-15074, 1986
 A:Title: Developmental and tissue-specific expression of alpha-1-microglobulin mRNA I
 A:Reference number: A25935; MUID:87033744
 A:Accession: A25935
 A:Molecule type: protein
 A:Residues: 141,'A',143-195 <KAS>
 R:Kido, H.; Yokogoshi, Y.; Katunuma, N.
 J. Biol. Chem. 263, 18104-18107, 1988
 A:Title: Kunitz-type protease inhibitor found in rat mast cells. Purification, proper
 A:Reference number: A31890; MUID:89053978
 A:Accession: A31890
 A:Molecule type: protein
 A:Residues: 283-301,'L',303-322,'N',324-329,'PK',332-333,'W',335-343 <KID>
 R:Singh, M.; Maruyama, M.; Yoshida, E.; Suni, H.; Mihar, H.
 Inflammation 15, 281-289, 1991
 A:Title: Acid-stable protease inhibitor in chronic phase of carrageenin-induced infla
 A:Reference number: A61633; MUID:92120777
 A:Accession: A61633
 A:Molecule type: protein
 A:Residues: 205-213,'X',215-229,'W',231-232,'K',234-238 <SUG>
 C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocali
 C:Keywords: chondroitin sulfate proteoglycan; chromoprotein; glycoprotein; plasma; se
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-202/Product: alpha-1-microglobulin #status predicted <A1M>
 F:34-187/Domain: lipocalin homology <LIP>
 F:205-349/Product: inter-alpha-trypsin inhibitor #status experimental <IAI>
 F:230-280/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:286-336/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 F:24/Binding site: carbohydrate (Thr) (covalent) #status predicted
 F:52/Cross-link: alpha-1-microglobulin-19 alpha complex chromophore (Cys) (interchain
 F:114,233/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:214/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental
 F:296/Inhibitory site: Arg (trypsin) #status predicted

Query Match 30.9%; Score 205; DB 2; Length 349;
 Best Local Similarity 39.8%; Pred. No. 2,7e-12;
 Matches 37; Conservative 14; Mismatches 40; Indels 2; Gaps 1;

OY 19 WHDPQGGCGMTFPARGCGARGFETYEACQACARCGDVCYLPAYOGR 78
 DB 246 YYYNGASMADETFYQGGCLNGNMFSEKECLQTRTA--ACNLPYQPCRAFMELWA 303
 OY 79 YSPLLQGCHEFYVGGCEGNGNMFHSRSCEDAC 111
 DB 304 FDAQKQKCIQFTYGGCKGNGNMFSEKECKEYC 336

RESULT 4
 HCHU
 alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor (validated) - human
 N:Alternate names: bikunin; complex-forming glycoprotein heterogeneous in charge (HC)
 rich protein
 N:Contains: alpha-1-microglobulin (protein HC); inter-alpha-trypsin inhibitor
 C:Species: Homo sapiens (man)
 C:Date: 15-Oct-1982 #sequence,revision 30-Jun-1987 #text,change 08-Dec-2000
 C:Accession: S13433; S10778; A93642; A90074; A90225; A90686; P04550; B39079; A61580;
 3217
 R:Veltr, H.; Gebhard, W.
 Biol. Chem. Hoppe-Seyler 371, 1185-1196, 1990
 A:Title: Structure of the human alpha(1)-microglobulin-bikunin gene.
 A:Reference number: S13433; MUID:91214554
 A:Accession: S13433
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-352 <VER1>
 A:Cross-references: EMBL:X54816; NID:q24475; PIDN:CAA3685.1; PID:9825614; EMBL:X5481
 R:Platani-Mehpour, M.; Bourguignon, J.; Sesboue, R.; Saller, J.P.; Devillard, T.; M
 Eur. J. Biochem. 191, 131-139, 1990
 A:Title: Structural analysis of the human inter-alpha-trypsin inhibitor light-chain g
 A:Reference number: S10778; MUID:90336621

A:Accession: S10778
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-202 <DIA>
 R:Kaunmeyer, J.F.; Polaszki, J.O.; Kotlick, M.P.
 Nucleic Acids Res. 14, 7839-7850, 1986
 A>Title: The mRNA for a proteinase inhibitor related to the HI-30 domain of Inter-alpha-
 A:Reference number: A93642; PMID:87040757
 A:Accession: A93642
 A:Molecule type: mRNA
 A:Residues: 1-352 <KAU>
 A:Cross-references: GB:X04494; NID:924478; PIDN:CAA28182.1; PID:g24479
 R:Lopez-Otin, C.; Grubb, A.O.; Mendez, E.
 Arch. Biochem. Biophys. 228, 544-554, 1984
 A>Title: The complete amino acid sequence of human complex-forming glycoprotein heteroge
 A:Reference number: A90074; PMID:84126849
 A:Accession: A90074
 A:Molecule type: protein
 A:Residues: 20-56,58-202 <LOP>
 A:Experimental source: Individual with tubular proteinuria
 A>Note: no evidence of sequence heterogeneity could be found, in spite of persistent het
 R:Takagi, T.; Takagi, K.; Kawai, T.
 Biochem. Biophys. Res. Commun. 98, 997-1001, 1981
 A>Title: Complete amino acid sequence of human alpha-1-microglobulin.
 A:Reference number: A90225; PMID:81184038
 A:Accession: A90225
 A:Molecule type: protein
 A:Residues: 20-47;58-136,138-141,'T',143-144,146-198 <TRK>
 A:Experimental source: pooled urine of patients with tubular proteinuria
 R:Reisner, P.; Hochstrasser, K.; Albrecht, G.J.; Lampert, K.; Salzer, J.P.
 Biol. Chem. Hoppe-Seyler 366, 479-483, 1985
 A>Title: Human inter-alpha-trypsin inhibitor: localization of the kunitz-type domains in
 A:Reference number: A90686; PMID:85225968
 A:Accession: A90686
 A:Molecule type: protein
 A:Residues: 206-290,'V',293-342,'E',344-350 <REI>
 R:Altman, F.; Lacour, B.; Strecker, G.; Parvy, P.; Druecke, T.; Daudon, M.
 Biochem. Biophys. Res. Commun. 191, 1156-1165, 1993
 A>Title: Molecular characteristics of uronic-acid-rich protein, a strong inhibitor of co
 A:Reference number: PNO450; PMID:93221481
 A:Accession: PNO450
 A:Molecule type: protein
 A:Residues: 206-214,'X' <ATML>
 R:Englind, J.J.; Salvesen, G.; Hefta, S.A.; Thøgersen, I.B.; Rutherford, S.; Pilz, S.V.
 J. Biol. Chem. 266, 747-751, 1991
 A>Title: Chondroitin 4-sulfate covalently cross-links the chains of the human blood prot
 A:Reference number: A39079; PMID:91093267
 A:Accession: B39079
 A:Molecule type: protein
 A:Residues: 206-225 <EMG1>
 R:Chirat, F.; Balduyck, M.; Mizon, C.; Laroui, S.; Sautiere, P.; Mizon, J.
 Int. J. Biochem. 23, 1201-1203, 1991
 A>Title: A chondroitin-sulfate chain is located on serine-10 of the urinary trypsin inh
 A:Reference number: A61580; PMID:92175157
 A:Accession: A61580
 A:Molecule type: protein
 A:Residues: 214,'X',216-222,'X' <CHT>
 R:McKeenhan, W.L.; Sakagami, Y.; Hoshi, H.; McKeenhan, K.A.
 J. Biol. Chem. 261, 5378-5383, 1986
 A>Title: Two apparent human endocytosomal cell growth factors from human hepatoma cells ar
 A:Reference number: A92583; PMID:86168278
 A:Accession: B25604
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-230,'X',232-239,'X',241-248,'XX',251-252,'X',254 <MCK>
 R:Englind, J.J.; Thøgersen, I.B.; Pilz, S.V.; Salvesen, G.
 J. Biol. Chem. 264, 15975-15981, 1989
 A>Title: Analysis of inter-alpha-trypsin inhibitor and a novel trypsin inhibitor, pre-
 A:Reference number: A92736; PMID:89380192
 A:Accession: C34245
 A:Molecule type: protein
 A:Residues: 206-225 <ENG2>
 R:Traboulsi, C.; Cortese, R.
 Nucleic Acids Res. 14, 6340, 1986

A>Title: Sequence of a full length cDNA coding for human protein HC (alpha-1-microglob
 A:Reference number: A25303; PMID:86312901
 A:Accession: A25303
 A:Molecule type: mRNA
 A:Residues: 1-218,'HW' <TRA>
 A>Note: this mRNA sequence appears to contain errors after residue 218
 R:Calero, M.; Escribano, J.; Grubb, A.; Mendez, E.
 J. Biol. Chem. 269, 384-389, 1994
 A>Title: Location of a novel type of interpolyptide chain linkage in the human prot
 A:Reference number: A53110; PMID:94103241
 A:Accession: A53110
 A:Molecule type: protein
 A:Residues: 45-57 <CAL>
 R:Vet, H.; Koegler, M.; Gebhard, W.
 FEBS Lett. 245, 137-140, 1989
 A>Title: The domain structure of the inhibitor subunit of human inter-alpha-trypsin 1
 A:Reference number: S03552; PMID:89171290
 A:Accession: S03552
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 206-352 <VBT2>
 R:Maikl, N.; Balduyck, M.; Maes, P.; Capon, C.; Mizon, C.; Han, K.K.; Tartar, A.; Fou
 Biol. Chem. Hoppe-Seyler 373, 1009-1018, 1992
 A>Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolat
 A:Reference number: S28928; PMID:93039735
 A:Accession: S28930
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-215 <MAL>
 R:Morelle, M.; Capon, C.; Balduyck, M.; Sautiere, P.; Kouach, M.; Michalski, C.; Fou
 Eur. J. Biochem. 221, 881-888, 1994
 A>Title: Chondroitin sulphate covalently cross-links the three polypeptide chains of
 A:Reference number: S43466; PMID:94229087
 A:Accession: S43466
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-221 <MOR>
 R:Kisilewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilek, J.
 Biochemistry 33, 7423-7429, 1994
 A>Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable co
 A:Reference number: A53642; PMID:94271799
 A:Accession: A53642
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-217 <WIS>
 R:Calero, M.; Mendez, E.; Garcia, E.
 Biochim. Biophys. Acta 1249, 91-99, 1995
 A>Title: Expression of the human complex-forming glycoprotein HC (alpha-1-microglobul
 A:Reference number: S55688; PMID:95284116
 A:Accession: S55688
 A:Molecule type: protein
 A:Residues: 20-24 <AL2>
 R:Bourguignon, J.; Diarra-Mehrput, M.; Seabone, R.; Frain, M.; Sala-Trepal, J.M.; Ma
 Biochem. Biophys. Res. Commun. 191, 1146-1153, 1985
 A>Title: Human inter-alpha-trypsin-inhibitor: characterization and partial nucleotide
 A:Reference number: I52208; PMID:86025577
 A:Accession: I52208
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 302-352 <BOU>
 A:Cross-references: GB:M11562; NID:9186587; PIDN:AA59194.1; PID:9307077
 R:Kojick, E.G.C.; van den Berg, M.; van der Linden, I.K.; Poort, S.R.; Cupers, R.; Be
 Biochem. J. 311, 753-759, 1995
 A>Title: Factor IX zutphen: a Cys(18) -> Arg mutation results in formation of a heter
 A:Reference number: S59509; PMID:96067589
 A:Accession: S59509
 A:Molecule type: protein
 A:Residues: 27-35,'Y',37 <MOJ>
 R:Altman, F.; Mizon, J.; Khan, S.R.
 Eur. J. Biochem. 236, 984-990, 1996
 A>Title: Identification of uronic-acid-rich protein as urinary bikunin, the light cha
 A:Reference number: S66434; PMID:96270753
 A:Accession: S66434

A:Molecule type: protein
 A:Residues: 206-214,'X',216-220 <ATM2>
 R:Akerstroem, B.; Bratt, T.; Enghild, J.J.
 FEBS Lett. 362, 50-54, 1995
 A:Title: Formation of the alpha(1)-microglobulin chromophore in mammalian and insect cell
 A:Reference number: S68728; MUID:95212582
 A:Accession: S68728
 A:Molecule type: protein
 A:Residues: 89-100 <AKS>
 R:Jessen, T.E.; Faarvang, K.L.; Ploug, M.
 FEBS Lett. 230, 195-200, 1988
 A:Title: Carbohydrate as covalent crosslink in human inter-alpha-trypsin inhibitor: a no
 A:Reference number: S02431; MUID:88167187
 A:Accession: S02431
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-217 <JES>
 R:Lopez, C.; Grubb, A.; Mender, E.
 FEBS Lett. 144, 349-353, 1982
 A:Title: Human protein HC displays variability in its carboxyl-terminal amino acid sequ
 A:Reference number: A91304
 A:Contents: annotation: variant of alpha-1-microglobulin
 A:Note: pooled urine samples contained two forms of this protein, both lacking 57-Lys an
 R:Hochstrasser, K.; Schonberger, O.L.; Rossmann, I.; Wachter, E.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 1357-1362, 1981
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-a
 by affinity chromatography.
 A:Reference number: A91698; MUID:82074265
 A:Contents: annotation: carbohydrate binding sites
 R:Morill, M.; Travis, J.
 Biol. Chem. Hoppe-Seyler 366, 19-21, 1985
 A:Title: The reactive site of human inter-alpha-trypsin inhibitor is in the amino-termi
 A:Reference number: A90682; MUID:85225940
 A:Contents: annotation: inhibitory site
 A:Note: in vitro, the first twelve residues of the amino end of the inhibitor appear to
 C:Comment: Alpha-1-microglobulin and inter-alpha-trypsin inhibitor are proteolytically
 C:Comment: Alpha-1-microglobulin occurs in many physiological fluids including plasma, u
 . It contains at least one brown-yellow chromophore.

Query Match 30.8%; Score 204; DB 1; Length 352;
 Best Local Similarity 39.4%; Pred. No. 3.4e-12;
 Matches 41; Conservative 12; Mismatches 45; Indels 6; Gaps 2;

QY 8 CTGTPSHLVLMHYDPQRCGCTFPARGCDGARGFEYEAQQAARGDGVLPVAVQ 67
 Db 240 CNGMTSRFY-----YNGTSMACETFPYGGCGNGNMFYEKECLQCRVVA--ACNLPVYR 293

QY 68 GPCRGMPRNAYSPILQOCHPFYVGGCGNGNMFHRSCEDEDAC 111
 Db 294 GPCRAFIOLMAFDVAKGKCVLPFYGGCGNGNMFYSEKREYEC 337

RESULT 5
 TIHOBI
 alpha-1-microglobulin/inter-alpha-trypsin inhibitor - horse (fragment)
 N:Alternate names: EI-14 (inhibitory fragment of ITI); ITI; trypsin inhibitor, E-UTI
 C:Species: Equus caballus (domestic horse)
 C:Date: 30-Jun-1987 #sequence,revision 04-Feb-2000 #text_change 05-May-2000
 R:Accession: A01210; A45653
 R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Reisinger, P.
 Biol. Chem. Hoppe-Seyler 366, 473-478, 1985
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-a
 A:Reference number: A90685; MUID:85225967
 A:Accession: A01210
 A:Molecule type: protein
 A:Residues: 3-125 <HOC>
 R:Veeravagavan, K.; Singh, K.; Wachter, E.; Hochstrasser, K.
 Biochem. Int. 26, 405-413, 1992
 A:Title: Characterization of a trypsin inhibitor from equine urine.
 A:Reference number: A45653; MUID:92328813
 A:Accession: A45653
 A:Status: preliminary
 A:Molecule type: protein

A:Residues: 1-12,'E',14-33 <VEE>
 A:Cross-references: PIDN:AAB22430.1; PID:9250858
 A:Experimental source: urine
 A:Note: sequence extracted from NCBI backbone (NCBI:107966)
 C:Comment: This inhibitory fragment, released from native ITI after limited proteolys
 First domain interacts weakly with PMN-granulocytic elastase and not at all with panc
 d elastase: The amino acid at position p2' (19-Met) appears to determine the specific
 C:Superfamily: protein HC, animal Kunitz-type proteinase inhibitor homology: lipocall
 C:Keywords: duplication; glycoprotein; plasma; serine proteinase inhibitor
 E:7-5//Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
 E:63-113//Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 E:7-5//16-40,32-53,63-113,72-96,88-109/Disulfide bonds: #status predicted
 E:17//Inhibitory site: leu (chymotrypsin, elastase) #status predicted
 E:26//Binding site: carbohydrate (Asn) (covalent) #status experimental
 E:73//Inhibitory site: Arg (trypsin) #status predicted

Query Match 30.2%; Score 200; DB 1; Length 125;
 Best Local Similarity 38.2%; Pred. No. 3.3e-12;
 Matches 39; Conservative 13; Mismatches 48; Indels 2; Gaps 1;

QY 10 GPTSPHVLVLMHYDPQRCGCTFPARGCDGARGFEYEAQQAARGDGVLPVAVQ 69
 Db 14 GPCLMISRYFNCTSMACETFPYGGCGNGNMFYSEKRECLQCRVVA--ACNLPVYGP 71

QY 70 CNGMPRNAYSPILQOCHPFYVGGCGNGNMFHRSCEDEDAC 111
 Db 72 CRAFTRLMAFDVAKGKCVLPFYGGCGNGNMFYSEKREYEC 113

RESULT 6
 TIHOBI
 alpha-1-microglobulin / inter-alpha-trypsin inhibitor precursor [validated] - bovine
 N:Alternate names: BI-14 (inhibitory fragment of ITI); bikunin; ITI
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 25-Feb-1985 #sequence,revision 04-Feb-2000 #text_change 18-Aug-2000
 R:Accession: S68149; A91717; A90685; S31219; A01209
 R:Lindqvist, A.; Akerstroem, B.
 Biochim. Biophys. Acta 1306, 98-106, 1996
 A:Title: Bovine alpha(1)-microglobulin/bikunin. Isolation and characterization of 11v
 A:Reference number: S68149; MUID:96201710
 A:Accession: S68149
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-352 <LIN>
 A:Cross-references: EMBL:U35642; NID:q1016297; PIDN:AAB07599.1; PID:q1016298
 R:Hochstrasser, K.; Wachter, E.
 Hoppe-Seyler's Z. Physiol. Chem. 364, 1679-1687, 1983
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte
 A:Reference number: A91717; MUID:84133807
 A:Accession: A91717
 A:Molecule type: protein
 A:Residues: 227-267,'L',269-273,'Q',275-297,'AE',300-329,'Q',331-345,'R',347-348 <HOC>
 R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Reisinger, P.
 Biol. Chem. Hoppe-Seyler 366, 473-478, 1985
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte
 A:Reference number: A90685; MUID:85225967
 A:Accession: A90685
 A:Molecule type: protein
 A:Residues: 347-349 <HOC2>
 R:Hochstrasser, K.; Albrecht, G.J.; Schonberger, O.L.; Wachter, E.
 Hoppe-Seyler's Z. Physiol. Chem. 364, 1689-1696, 1983
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte
 A:Reference number: A91718; MUID:84133808
 A:Contents: annotation: reactive sites
 R:Castillo, G.M.; Templeton, D.M.
 FEBS Lett. 318, 292-296, 1993
 A:Title: Subunit structure of bovine EBF (extracellular matrix stabilizing factor(s))
 A:Reference number: S31219; MUID:93178646
 A:Accession: S31219
 A:Status: preliminary
 A:Molecule type: protein

A:Title: cDNA sequence of rabbit lipoprotein-associated coagulation inhibitor.
 A:Reference number: S12143; MUID:91057146
 A:Accession: S12143
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-300 <RES>
 A:Cross-references: EMBL:X54708; NID:91612; PIDN:CAA38515.1; PID:91613
 R:Goldbun, P.; Grabb, J.W.; Buonassisi, V.
 J. Cell. Physiol. 148, 320-326, 1991
 A:Title: Enhanced inhibition of tissue factor by the extended form of an endothelial cell
 A:Reference number: A61373; MUID:91349227
 A:Accession: A61373
 A:Molecule type: protein
 A:Residues: 25-33, 'X', 35-46 <COL>
 A:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor
 C:Keywords: anticoagulant; glycoprotein
 F:50-100/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:121-171/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 F:213-263/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>

Query Match 28.7%; Score 190.5; DB 2; Length 300;
 Best Local Similarity 34.8%; Pred. No. 5.6e-11;
 Matches 40; Conservative 15; Mismatches 47; Indels 13; Gaps 1;

Oy 10 GPTSPHLVLMHYDPQGGCMTPPARCGDGAARGFETVACQOACAR-----G 56
 Db 57 GPCRAYIKRFEPFNILAHQCEFIYGGCEGNENRFSLECKEKARQPKMTKLTFFQKG 116
 Oy 57 PGDAGVLPVAVGPGRGWPRRAYSPLIQCPHYVYGGCEGNENRFSLECKEKARQPKMTKLTFFQKG 111
 Db 117 KPDCFLIEDPFGICRGYITRIFVNNQSKQCFRFRYGGLGLNLFNFSLECKKNTC 171

RESULT 11
 JG0185
 hepatocyte growth factor activator inhibitor type 2 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
 C:Accession: JG0185
 R:Itch, H.; Kataoka, H.; Hamasuna, R.; Kitamura, N.; Kono, M.
 Biochem. Biophys. Res. Commun. 255, 740-748, 1999
 A:Title: Hepatocyte growth factor activator inhibitor type 2 lacking the first kunitz-ty
 A:Reference number: JG0185; MUID:99160423
 A:Accession: JG0185
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-252 <IT0>
 A:Cross-references: GB:AF099016
 C:Superfamily: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:133-183/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>

Query Match 28.4%; Score 188.5; DB 2; Length 252;
 Best Local Similarity 28.5%; Pred. No. 7.5e-11;
 Matches 41; Conservative 12; Mismatches 54; Indels 37; Gaps 2;

Oy 5 VOACTGPTSHLVLMHYDPQGGCMTPPARCGDGAARGFETVACQOACA----- 54
 Db 40 VSKVVGCRASIPRMWYNTIDGSCOPVYGGCEGNENRFSLECKEKARQPKMTKLTFFQKG 99
 Oy 55 ---RGPGA-----CVLPVAVGPGRGWPRRAYSPLIQCH 87
 Db 100 ARNRNGADSSVLPKQSAEDLSAEFTNEXCVPRAYVGPCRAAPRMYDTEKNSCI 159

Oy 88 PAVYGGCEGNENRFSLECKEKARQPKMTKLTFFQKG 111
 Db 160 SFYVGGCRGNKNSYLSQEGACMQHC 183

RESULT 12
 TTHUGK
 tissue factor pathway inhibitor precursor [validated] - human

N:Alternate names: extrinsic pathway inhibitor; lipoprotein-associated coagulation in
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Dec-2000
 C:Accession: A23712; A39176; A28650; A60433; S13034; A34315; A38294; S03903
 R:Girard, T.J.; Eddy, R.; Wesselschmidt, R.L.; Macphail, L.A.; Likert, K.M.; Byers, M
 J. Biol. Chem. 266, 5036-5041, 1991
 A:Title: Structure of the human lipoprotein-associated coagulation inhibitor gene. In
 A:Reference number: A23712; MUID:91161593
 A:Accession: A23712
 A:Molecule type: DNA
 A:Residues: 1-304 <GIR>
 A:Cross-references: GB:M59493; GB:M59499; NID:9187204; PIDN:AAA59526.1; PID:9187206
 R:van der Logt, C.P.E.; Reitsma, P.H.; Bertina, R.M.
 Biochemistry 30, 1571-1577, 1991
 A:Title: Intron-exon organization of the human gene coding for the lipoprotein-associ
 A:Reference number: A39176; MUID:91129227
 A:Accession: A39176
 A:Molecule type: DNA
 A:Residues: 1-304 <VAN>
 A:Cross-references: GB:M58650; GB:J05312; NID:9186827; PIDN:AAA59480.1; PID:9186829
 R:Wun, T.C.; Kretzmer, K.K.; Girard, T.J.; Miletich, J.P.; Broze Jr., G.J.
 J. Biol. Chem. 263, 6001-6004, 1988
 A:Title: Cloning and characterization of a cDNA coding for the lipoprotein-associated
 A:Reference number: A28650; MUID:88198127
 A:Accession: A28650
 A:Molecule type: mRNA
 A:Residues: 1-304 <MUN>
 A:Cross-references: GB:J03325; NID:9180545; PIDN:AAA52022.1; PID:9180546
 A:Note: part of this sequence, including the amino end of the mature protein, was con
 R:Girard, T.J.; Warren, L.A.; Novotny, W.F.; Bejcek, B.E.; Miletich, J.P.; Broze Jr.,
 Thromb. Res. 55, 37-50, 1989
 A:Title: Identification of the 1.4 KB and 4.0 KB messages for the lipoprotein associa
 A:Reference number: A60433; MUID:89388722
 A:Accession: A60433
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-304 <G12>
 A:Experimental source: endothelial cells
 A:Accession: B60433
 A:Molecule type: protein
 A:Residues: 'XX', 31-53, 'X', 55-56 <G13>
 A:Experimental source: recombinant material from mouse C137 cells
 R:Girard, T.J.; McCourt, D.; Novotny, W.F.; Macphail, L.A.; Likert, K.M.; Broze Jr.,
 Biochem. J. 270, 621-625, 1990
 A:Title: Endogenous phosphorylation of the lipoprotein-associated coagulation inhibit
 A:Reference number: S13034; MUID:91054349
 A:Accession: S13034
 A:Molecule type: protein
 A:Residues: 29-35 <G14>
 R:Novotny, W.F.; Girard, T.J.; Miletich, J.P.; Broze Jr., G.J.
 J. Biol. Chem. 264, 18832-18837, 1989
 A:Title: Purification and characterization of the lipoprotein-associated coagulation
 A:Reference number: A34315; MUID:90036996
 A:Accession: A34315
 A:Molecule type: protein
 A:Residues: 'XX', 31-33, 'L', 35-50 <NOV>
 A:Experimental source: plasma
 R:Pedersen, A.H.; Nordand, O.; Norris, F.; Wiberg, F.C.; Christensen, P.M.; Moeller,
 J. Biol. Chem. 265, 16786-16793, 1990
 A:Title: Recombinant human extrinsic pathway inhibitor. Production, isolation, and ch
 A:Reference number: A38294; MUID:91009092
 A:Accession: A38294
 A:Molecule type: protein
 A:Residues: 29-41 <PED>
 R:Girard, T.J.; Warren, L.A.; Novotny, W.F.; Likert, K.M.; Brown, S.G.; Miletich, J.P
 Nature 338, 518-520, 1989
 A:Title: Functional significance of the Kunitz-type inhibitory domains of lipoprotein
 A:Reference number: S03903; MUID:89181950
 A:Contents: annotation; site-directed mutagenesis
 C:Comment: The first Kunitz-type domain binds the factor VIIa/tissue factor complex;
 C:Genetics:
 A:Gene: GDB:TFPI
 A:Cross-references: GDB:127364; OMIM:152310

A:Map position: 2q32-2q32
 A:Introns: 41/1; 107/1; 120/1; 179/1; 210/1; 270/1
 C:Function:
 A:Description: regulates clotting by factor Xa-dependent inhibition of the coagulation factor X
 A:Pathway: blood coagulation
 C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor
 C:Keywords: anticoagulant; blood coagulation; duplication; glycoprotein; heparin binding
 F:1-28/Domain: signal sequence status predicted <Sig>
 F:29-304/Product: tissue factor pathway inhibitor #status experimental <MAT>
 F:54-104/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:125-175/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 F:217-267/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>
 F:284-289/Region: heparin binding #status predicted
 F:30/Binding site: phosphate (Ser) (covalent) #status experimental
 F:54-104,63-87,79-100,125-175,134-158,150-171,217-267,226-250,242-263/Disulfide bonds: #
 F:64/Inhibitory site: Lys (coagulation factor VII/tissue factor complex) #status experim
 F:135/Inhibitory site: Arg (coagulation factor X) #status experimental
 F:145,195,256/Binding site: carbonylate (Asn) (covalent) #status predicted
 F:227/Inhibitory site: Arg (unidentified proteinase) #status predicted

Query Match 28.4%; Score 188; DB 1; Length 304;
 Best Local Similarity 29.6%; Pred. No. 9,9e-11;
 Matches 42; Conservative 14; Mismatches 48; Indels 38; Gaps 2;

OY 4 DVQACTPTSPHLVLMHYDPORGCMTFPARGCDGARGFEYEAQOACARG-----58
 DB 130 DEICNG-----YTRTFYNNQKQCFKFGGCLGMNNEFTLEBCKNCEDLNGRVD 185
 OY 59 -----DACVLPVPGPCRGWEPRMAYSPLLQOCHP 89
 DB 186 NYGTQLAVANNSTPQSTKVPSPFEEFHP-SWCLAPADRLCRANERFYNSVIGCRP 245
 OY 90 VYGGCEGNGNFRHRESCEDAC 111
 DB 246 KYSGCGGNENFTSKRECLRAC 267

RESULT 13
 TIRTK
 tissue factor pathway inhibitor precursor - rat
 N:Alternate names: extrinsic pathway inhibitor; lipoprotein-associated coagulation inhib
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jun-2000
 C:Accession: JX0213
 R:Enzymol., K.; Emi, M.; Mukai, T.; Kato, H.
 J. Biochem. 111, 681-687, 1992
 A:Title: cDNA cloning and expression of rat tissue factor pathway inhibitor (TFPI).
 A:Reference number: JX0213; MUID:92248361
 A:Accession: JX0213
 A:Molecule type: mRNA
 A:Residues: 1-302 <END>
 A:Cross-references: DDBJ:D10926; NID:g220916; PIDN:BA01724.1; PID:g220917
 A:Experimental source: liver
 C:Comment: This serine proteinase inhibitor regulates clotting by factor Xa-dependent in
 C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor
 C:Keywords: anticoagulant; blood coagulation; duplication; glycoprotein; heparin binding
 F:1-28/Domain: signal sequence status predicted <Sig>
 F:29-302/Product: tissue factor pathway inhibitor #status predicted <MAT>
 F:53-103/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:124-174/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 F:222-272/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>
 F:286-291/Region: heparin binding #status predicted
 F:53-103,62-86,78-99,124-174,133-157,149-170,222-272,231-255,247-268/Disulfide bonds: #
 F:63/Inhibitory site: Lys (coagulation factor VII/tissue factor complex) #status predict
 F:134/Inhibitory site: Arg (coagulation factor X) #status predicted
 F:144,251,261/Binding site: carbonylate (Asn) (covalent) #status predicted
 F:232/Inhibitory site: Lys (unidentified proteinase) #status predicted

Query Match 27.2%; Score 180.5; DB 1; Length 302;
 Best Local Similarity 31.7%; Pred. No. 5e-10;

Matches 38; Conservative 16; Mismatches 53; Indels 13; Gaps 1;

OY 5 VOACTPTSPHLVLMHYDPORGCMTFPARGCDGARGFEYEAQOACARG-----56
 DB 55 MKRDEPCRAMTRSYTFNMNSHOCSEFYGGCGNKNRDTLECKRTCPGKTKTKT 114
 OY 57 -----PGDACVLPVPGPCRGWEPRMAYSPLLQOCHPFGYGGCEGNGNFRHRESCEDAC 111
 DB 115 TSGAEKPDCELEEDDGICRGFTKRFYNNQSKQCFKFGGCLGMNNEFTLEBCKRNC 174

RESULT 14
 JC2264
 tissue factor pathway inhibitor precursor - rhesus macaque
 N:Alternate names: extrinsic pathway inhibitor; lipoprotein-associated coagulation in
 C:Species: Macaca mulatta (rhesus macaque)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: JC2264
 R:Kamel, S.; Kamikubo, Y.; Hamuro, T.; Fujimoto, H.; Ishihara, M.; Yonemura, H.; Miya
 J. Biochem. 115, 708-714, 1994
 A:Title: Amino acid sequence and inhibitory activity of rhesus monkey tissue factor p
 A:Reference number: JC2264; MUID:94375417
 A:Accession: JC2264
 A:Molecule type: mRNA
 A:Residues: 1-304 <KAM>
 A:Cross-references: GB:S73337; NID:9685016; PIDN:AA831955.1; PID:9685017
 A:Experimental source: liver
 C:Comment: This protein inhibits the activities of factor Xa and tissue factor-factor
 C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibit
 C:Keywords: anticoagulant; glycoprotein; serine proteinase inhibitor
 F:1-28/Domain: signal sequence status predicted <Sig>
 F:29-304/Product: tissue factor pathway inhibitor #status predicted <MAT>
 F:54-104/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
 F:125-175/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
 F:217-267/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>
 F:54-104,63-87,79-100,125-175,134-158,150-171,217-267,226-250,242-263/Disulfide bonds
 F:64/Inhibitory site: Lys (coagulation factor VII/tissue factor complex) #status pred
 F:135/Inhibitory site: Arg (coagulation factor X) #status predicted
 F:145,195,256/Binding site: carbonylate (Asn) (covalent) #status predicted
 F:227/Inhibitory site: Arg (unidentified proteinase) #status predicted

Query Match 26.8%; Score 178; DB 1; Length 304;
 Best Local Similarity 29.4%; Pred. No. 8.7e-10;
 Matches 42; Conservative 13; Mismatches 48; Indels 40; Gaps 3;

OY 4 DVQACTPTSPHLVLMHYDPORGCMTFPARGCDGARGFEYEAQOACARG-----54
 DB 130 DEICNG-----YTRTFYNNQSKQCFKFGGCLGMNNEFTLEBCKNCEDLNGRVD 185
 OY 55 -----RGPDACVLPVPGPCRGWEPRMAYSPLLQOCHP 88
 DB 186 NYGTQLAVANNSTPQSTKVPSPFEEFHP-SWCLAPADRLCRANERFYNSVIGCRP 244
 OY 89 FVYGGCEGNGNFRHRESCEDAC 111
 DB 245 FKYSGGGNENFTSKRECLRAC 267

RESULT 15
 T26063
 hypothetical protein W01F3.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26063
 R:Cummings, P.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z20145
 A:Accession: T26063
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2225 <WIL>
 A:Cross-references: EMBL:292815; PIDN:CAB07294.1; GSPDB:GN00023; CESP:W01F3.3

A:Experimental source: clone W01F3

C:Genetics:

A:Gene: CESP:W01F3.3

A:Map position: 5

A:Introns: 33/1; 56/1; 100/1; 142/3; 271/3; 451/1; 525/3; 774/1; 1093/1; 1178/1; 1221/1;

Query Match

26.5%; Score 175.5; DB 2; Length 2225;

Best Local Similarity 31.0%; Pred. No. 8.8e-09;

Matches 35; Conservative 9; Mismatches 56; Indels 13; Gaps 2;

QY 8 CTGPTSPHLVLMHYDPQRGCGMTPPARGCDGARGFETYEACQACARP-----G 58

Db 786 CRG----QFVWVFDEDEKKNCDVFTYTGCGGNGNNFASKKECMALCHKPEPTPSATPDES 841

QY 59 DACVLPAYOGPCRGWEPFRMAISPLIOCHFPFVYGGCEGNGNMFHSRSCEDAC 111

Db 842 QVCSNDVDAGBCNGVFERFAFDAEADODCRATFTYGGCGGNGNMFATMOECRSRC 894

Search completed: February 26, 2002, 01:29:42
Job time: 421 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:07 : Search time 69.26 Seconds
(without alignments)
36.065 Million cell updates/sec

Title: US-09-819-136-2_COPY_299_409
Perfect score: 663
Sequence: 1 CLDPVQACCTPTSPHLVLMH.....GGCGNGNNFHSRSCEDAC 111

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	204	30.8	143	2	US-08-422-333-10
2	204	30.8	143	6	5223482-20
3	204	30.8	144	6	5187153-18
4	204	30.8	147	1	US-08-358-160-72
5	201	30.3	123	6	5466783-32
6	200	30.2	123	6	5466783-21
7	200	30.2	127	6	5466783-24
8	199	30.0	122	2	US-08-422-333-12
9	199	30.0	122	6	5187153-20
10	199	30.0	122	6	5220013-23
11	198	29.9	122	6	5223482-32
12	192.5	28.0	122	6	5466783-23
13	190.5	28.7	125	1	US-08-685-660A-7
14	190.5	28.7	252	2	US-08-974-196-7
15	190.5	28.7	252	4	US-09-071-709-10
16	190.5	28.7	252	4	US-09-013-896A-2
17	189.5	28.6	213	6	5466783-25
18	188	28.4	276	1	US-07-828-920A-1
19	188	28.4	276	1	US-08-437-841-9
20	188	28.4	276	1	US-08-286-521-9
21	188	28.4	276	1	US-08-436-175-9
22	188	28.4	276	2	US-08-796-850-11
23	188	28.4	276	3	US-08-854-764-3
24	188	28.4	276	4	US-08-943-682-9
25	188	28.4	276	4	PCT-US95-09377-3
26	188	28.4	276	5	PCT-US95-09464-9
27	188	28.4	277	1	US-07-844-297-1

28	188	28.4	304	1	US-08-026-145-2	Sequence 2, Appl1
29	188	28.4	304	1	US-08-446-646-9	Sequence 9, Appl1
30	188	28.4	304	1	US-08-676-125A-18	Sequence 18, Appl1
31	188	28.4	304	2	US-09-136-012A-18	Sequence 1, Appl1
32	188	28.4	304	3	US-08-676-124-1	Sequence 25, Appl1
33	188	28.4	304	3	US-08-208-264A-25	Sequence 1, Appl1
34	188	28.4	304	3	US-09-414-878-1	Sequence 1, Appl1
35	188	28.4	304	3	US-09-240-136-1	Sequence 2, Appl1
36	188	28.4	304	4	US-09-054-782-2	Patent No. 5466783
37	188	28.4	304	6	5466783-2	Sequence 2, Appl1
38	188	28.4	352	3	US-08-854-764-2	Sequence 2, Appl1
39	188	28.4	352	5	PCT-US95-09377-2	Sequence 2, Appl1
40	174.5	26.3	161	1	US-08-437-841-19	Sequence 19, Appl1
41	174.5	26.3	161	1	US-08-286-521-19	Sequence 19, Appl1
42	174.5	26.3	161	1	US-08-436-175-19	Sequence 19, Appl1
43	174.5	26.3	161	4	US-08-943-682-19	Sequence 19, Appl1
44	174.5	26.3	161	5	PCT-US95-09464-19	Sequence 19, Appl1
45	174	26.2	213	2	US-08-796-850-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-422-333-10
; Sequence 10, Application US/08422333
; Patent No. 5912410
; GENERAL INFORMATION:
; APPLICANT: CORDELL, Barbara L.
; TITLE OF INVENTION: TRANSGENIC NON-HUMAN MAMMAL DISPLAYING
; TITLE OF INVENTION: THE AMYLOID-FORMING PATHOLOGY OF ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scios, Inc.
; STREET: 2450 Bayshore Parkway
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,333
; FILING DATE: 13-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Shearer, Peter R.
; REGISTRATION NUMBER: 28,117
; REFERENCE/DOCKET NUMBER: 21900-28048.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 966-1550
; TELEFAX: (415) 968-2438
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 143 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-422-333-10

Query Match 30.8%; Score 204; DB 2; Length 143;
Best Local Similarity 38.5%; Pred. No. 2.5e-13;
Matches 40; Conservative 13; Mismatches 45; Indels 6; Gaps 2;

QY 8 CCGPTSPHLVLMHYDDQRCGCMFFPARCGDGAARGETYEACQQAARGDCAVLPAYQ 67
DB 35 CCGMTSRYPF---YNGTSMACETFFQYGGCGMGNGNNFVTEKECLQTRTYA--ACNLPVIR 88
QY 68 GCGRGWEPRAWYSPLLQCHPEFYVGGCEGNGNMFHSRSCEDAC 111

Db 89 GPCRAFIOLMAFDAYKGVLPFYGGCGGNGNKFYSERECREYC 132

RESULT 2
5223482-20

Patent No. 5223482
APPLICANT: SCHILLING, JAMES W.; PONTE, PHYLLIS A.; CORDELL,
BARBARA
TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
NUMBER OF SEQUENCES: 34

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/361,912

FILING DATE: 06-JUN-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 359,911

FILING DATE: 12-MAY-1989

APPLICATION NUMBER: 87,002

FILING DATE: 18-AUG-1987

APPLICATION NUMBER: 8,810

FILING DATE: 30-JAN-1987

APPLICATION NUMBER: 948,376

FILING DATE: 31-DEC-1986

APPLICATION NUMBER: 932,193

FILING DATE: 17-NOV-1986

SEQ ID NO: 20:

LENGTH: 143

5223482-20

Query Match 30.8%; Score 204; DB 6; Length 143;
Best Local Similarity 38.5%; Pred. No. 2.5e-13;

Matches 40; Conservative 13; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVIMHYDPQRCGCMTPFARGCCGANGFEYECQACARGPEDACVLPVAVQ 67

Db 35 CMGMSIRFY----YNGTSMACETFOYGGCMGNGNMFYERKECLOTCTRTVA--ACNLPIVR 88

OY 68 GPCGWEPBRMAYSPDLOQCHPEYVYGCGCGNGNMFHRSRSCEDAC 111

Db 89 GPCRAFIOLMAFDAYKGVLPFYGGCGGNGNKFYSERECREYC 132

RESULT 3
5187153-18

Patent No. 5187153

APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S

AMYLOID POLYPEPTIDE DERIVATIVES
NUMBER OF SEQUENCES: 33

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/502,273

FILING DATE: 29-MAR-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 361,912

FILING DATE: 06-JUN-1989

APPLICATION NUMBER: 359,911

FILING DATE: 12-MAY-1989

APPLICATION NUMBER: 87,002

FILING DATE: 18-AUG-1987

APPLICATION NUMBER: 8,810

FILING DATE: 30-JAN-1987

APPLICATION NUMBER: 948,376

FILING DATE: 31-DEC-1986

APPLICATION NUMBER: 932,193

FILING DATE: 17-NOV-1986

SEQ ID NO: 18:

LENGTH: 144

5187153-18

Query Match 30.8%; Score 204; DB 6; Length 144;

Best Local Similarity 38.5%; Pred. No. 2.5e-13;
Matches 40; Conservative 13; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVIMHYDPQRCGCMTPFARGCCGANGFEYECQACARGPEDACVLPVAVQ 67

Db 35 CMGMSIRFY----YNGTSMACETFOYGGCMGNGNMFYERKECLOTCTRTVA--ACNLPIVR 88

OY 68 GPCGWEPBRMAYSPDLOQCHPEYVYGCGCGNGNMFHRSRSCEDAC 111

Db 89 GPCRAFIOLMAFDAYKGVLPFYGGCGGNGNKFYSERECREYC 132

RESULT 4
US-08-358-160-72

Sequence 72; Application US/08358160

Patent No. 5663143

GENERAL INFORMATION:

APPLICANT: LEY, Arthur C.

APPLICANT: LADNER, Robert C.

APPLICANT: GUTERMAN, Sonia K.

APPLICANT: ROBERTS, Bruce L.

APPLICANT: MARKLAND, William

APPLICANT: KENT, Rachel B.

TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ

TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE

NUMBER OF SEQUENCES: 234

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROMDY AND NEWMARK

STREET: 419 Seventh Street, N.W. Suite 300

CITY: Washington

STATE: District of Columbia

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/358,160

FILING DATE: 16-DEC-1994

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/133,031

FILING DATE: 13-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/009,319

FILING DATE: 26-JAN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/664,989

FILING DATE: 01-MAR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/487,063

FILING DATE: 02-MAR-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/240,160

FILING DATE: 02-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Cooper, Iver P.

REGISTRATION NUMBER: 28,005

REFERENCE/DOCKET NUMBER: LEY=1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 147 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-358-160-72


```

? NAME: Shearer, Peter R.
? REGISTRATION NUMBER: 28, 117
? REFERENCE/DOCKET NUMBER: 21900-28048.00
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 966-1550
? TELEFAX: (415) 968-2438
? INFORMATION FOR SEQ ID NO: 12:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 122 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
US-08-422-333-12

```

Query Match	30.0%	Score 199	DB 2	Length 122
Best Local Similarity	38.2%	Pred. No. 6.7e-13		
Matches	39	Conservative	15	Mismatches 46; Indels 2; Gaps 1

Dy 70 CRRGPRRAYSPILDOCHPFYGGCGEGGNFFHSRSCEDAC 111
 || : : : : || : : : : || : : : : || : : : :
 70 CRAFIQIMAFDAKGVCRFESYGCKGNGNFYSOKCKKEVC 111

RESULT 9
5187153-20
Patent No. 5187153
APPLICANT: CONDELL, BARBARA SCHILLING, JAMES W.; KATUNDMA, NOBOHIKO
TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER S
AMYLROID POLYPEPTIDE DERIVATIVES
NUMBER OF SEQUENCES: 33
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/502, 273
FILING DATE: 29-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 361, 912
FILING DATE: 06-JUN-1989
APPLICATION NUMBER: 359, 911
FILING DATE: 12-MAY-1989
APPLICATION NUMBER: 87, 002
FILING DATE: 18-AUG-1987
APPLICATION NUMBER: 8, 810
FILING DATE: 30-JAN-1987
APPLICATION NUMBER: 948, 376
FILING DATE: 31-DEC-1986
APPLICATION NUMBER: 932, 193
FILING DATE: 17-NOV-1986
SEQ ID NO: 20:
LENGTH: 122
5187153-20

Query Match	30.0%;	Score 199;	DB 6;	Length 122;
Best Local Similarity	38.2%;	Pred. No. 6.7e-13;		
Matches 39; Conservative	15;	Mismatches 46;	Indels 2;	Gaps 1;

QY 10 GPSPLVLMVHDPORGCGTTPAGCCGCGARGETEACGACGACRGGPACTVAVGSP 69
Db 12 GPGDLGFRKRYFNIGSMACEFFLYGCGCGNINLNLSSQELQTOR--TYEACNIPVGGP 69
QY 70 CRGWEPRRAYSPLLQCHPEFYGGCGEGNGNNFHSRSCEDAC 111
Db 70 CRAFIQLMAFDVAVKCKVRSFYGGCKGNGNNFYQSKKEVC 111

RESULT 10
5220013-23
;Patent No. 5220013
; APPLICANT: PONTE, PHYLLIS A.;CORDELL, BARBARA

```

1  TITLE OF INVENTION:  DNA SEQUENCE USEFUL FOR THE DETECTION
2  OF ALZHEIMER'S DISEASE
3  NUMBER OF SEQUENCES: 30
4  CURRENT APPLICATION DATA:
5  APPLICATION NUMBER:  US/07/444,118
6  FILING DATE:  30-NOV-1985
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER:  87, 002
9  FILING DATE:  18-AUG-1987
10 APPLICATION NUMBER:  8, 810
11 FILING DATE:  30-JAN-1987
12 APPLICATION NUMBER:  948, 376
13 FILING DATE:  31-DEC-1986
14 APPLICATION NUMBER:  932, 193
15 FILING DATE:  17-NOV-1986
16
17 SEQ ID NO:23:
18 LENGTH: 122
19
20 1220013-23

```

Query Match	30.0%;	Score 199;	DB 6;	Length 122;
Best Local Similarity	38.2%;	Pred. No. 6.7e-13;		
Matches 39;	Conservative 15;	Mismatches 46;	Indels 2;	Gaps 1

QY 10 GPTSPILVLMHNDPQGGCMETPARCGCGARGFEITYEACCOACARGPGACVLPANVG 65

Db 12 GPCLTGFKNRYFNGTSMACETFLYGCGAGNLNLFSLQKECLQICR--TVACNLPITVGG 69

QY 70 CRGCRERRRAYSPFLDQCCHFFVYGGCGEGCGNNFHSRSECDAC 111

Db 70 CRATIDLMFADAVKGCYARESTGCGCKGNGNFFYSQKECKEC 111

RESULT 11
 5223482-22
 ; Patent No. 5223482
 ; APPLICANT: SCHILLING, JAMES W.; PONTE, PHYLLIS A.; CORDELL,
 ; BARBARA
 ; TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
 ; INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
 ; NUMBER OF SEQUENCES: 34
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/361,912
 ; FILING DATE: 06-JUN-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 359,911
 ; FILING DATE: 12-MAY-1989
 ; APPLICATION NUMBER: 87,002
 ; FILING DATE: 18-AUG-1987
 ; APPLICATION NUMBER: 8,810
 ; FILING DATE: 30-JAN-1987
 ; APPLICATION NUMBER: 948,376
 ; FILING DATE: 31-DEC-1986
 ; APPLICATION NUMBER: 932,193
 ; FILING DATE: 17-NOV-1986
 ; SEQ ID NO:22:
 ; LENGTH: 122
 5223482-22

Query Match	29.94;	Score 198;	DB 6;	Length 122;
Best Local Similarity	39.84;	Pred. No. 8.4e-13;		
Matches 37; Conservative	15;	Mismatches 39;	Indels 2;	Gaps 1

[illegible]

This Page Blank (uspto)


```

XX PR 03-SEP-1998; 98US-0148092.
XX PA (ZYMO ) ZYMOGENETICS INC.
XX PI Conklin DC;
XX MP1: 2000-256985/22.
DR N-FSDB; AAZ99802, AAZ99803.
PT Novel proteinase inhibitor zkun6 comprising a knitz domain, useful in
PT the treatment or prevention of conditions associated with excessive
PS proteinase activity -
XX Claim 4; Page 40; 48pp; English.
CC The present sequence represents a serine protease inhibitor containing
CC a knitz domain, designated ZKUN6. The zkun6 polypeptides are used in
CC the treatment and prevention of conditions associated with excessive
CC proteinase activity. The conditions include acute pancreatitis,
CC cardiopulmonary bypass-induced pulmonary injury, allergy- induced
CC protease release, deep vein thrombosis, myocardial infarction,
CC shock, hyperfibrinolytic haemorrhage, emphysema, rheumatoid arthritis,
CC adult respiratory distress syndrome, chronic inflammatory bowel disease,
CC psoriasis, inflammatory conditions, platelet function, organ
CC preservation, and wound healing. They are also useful in the treatment
CC of conditions arising from haemostasis imbalance, including acquired
CC coagulopathies, primary fibrinolysis and fibrinolysis due to cirrhosis,
CC and complications of high-dose thrombolytic therapy. zkun6 polypeptides
CC may also be used for inhibiting blood coagulation in mammals, and for
CC blockade of protolytic tissue degradation. The zkun6 polynucleotides
CC are used in gene therapy to treat the above diseases. Transgenic animals,
CC engineered to express zkun6, and knockout animals with an absence of
CC zkun6 function, are used to study the zkun6 gene and the encoded
CC protein. They are useful for investigating the role of zkun6 polypeptides
CC in early development.
SQ Sequence 59 AA;
QY Query Match 51.3%; Score 340; DB 21; Length 59;
Best Local Similarity 100.0%; Pred. No. 8.3e-28;
Db 1 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0
56 GPGDNCVLPAYOGPGRGWEPRMAYSPLLQQCHPRFYGGCEGMGNFHSRSCEDAC 111
1 gpqdadcvlpavpgprgwpewrwaysplllqqchpfyggcegmgnfhsrcsedac 56
RESULT 2
ID AAR62523
AA R62523 standard; peptide; 560 AA.
AC AAR62523;
DT 06-JUN-1995 (first entry)
DE Hookworm anticoagulant.
EE Hookworm anticoagulant.
KW Hookworm; anticoagulant; serine protease-inhibitor; blood loss;
KW vaccine; vascular disease therapy.
OS Ancyllostoma caninum.
PN WO9425000-A.
XX 10-NOV-1994.
XX 29-APR-1994; 94WO-US04707.
XX 30-APR-1993; 93US-0055988.
XX (UYIA ) UNIV YALE.
```

[illegible]

XX Disclousre; Fig 8A; 72pp; English.

PS
XX
CC This invention describes novel transgenic mice expressing proteins
CC related to the pathology of Alzheimer's disease and which provide models
CC for studying potentially therapeutic compounds. The transgenic mice
CC contain a DNA sequence encoding a beta-amyloid precursor protein (APP)
CC and a nerve tissue specific promoter operably linked to the beta-APP
CC allowing its expression to form beta-amyloid protein deposits in the
CC animal's brain. The transgenic mouse is useful for elucidating the
CC molecular mechanisms involved in the synthesis of and, more importantly,
CC inhibiting the synthesis and deposition of beta-amyloid proteins (most
CC importantly in the brain where plaque formation is associated with
CC Alzheimer's disease) by inhibiting production and/or increasing cleavage
CC after production. The transgenic animals provide useful models for
CC studying the in vivo relationships of the proteins to each other and to
CC other compounds being tested for their usefulness in treating Alzheimer's
CC disease.

XX
SQ Sequence 143 AA:

Query Match 30.8%; Score 204; DB 20; Length 143;
Best Local Similarity 38.5%; Pred. No. 1.7e-13;
Matches 40; Conservative 13; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVLMHYDPQRCGCMTFPARCGDGAARGFTEYACQACARGPGDVCVPAVQ 67
DB 35 cmgmstsyf-----yngtmacetfgygcmgngnftvtekeclqctrtva--acnlpivr 88
OY 68 GRCRGHEPRMAYSPILQOCHPFYVGGCEGNGNMFHRSCECDAC 111
DB 89 gpcraflqlwaIdavkvcvllfpygqcgngnkfysekecreyc 132

RESULT 4
AAR92237
ID AAR92237 standard; protein; 147 AA.

XX
AC AAR92237;
XX
DT 27-SEP-1996 (first entry)
XX
DE Human wild-type urinary trypsin inhibitor.
XX
XX
KW UTI; kunitz domain; urinary trypsin inhibitor; elastase inhibitor;
KW recombinant protein production; yeast host cell; Pichia;
KW site-directed mutagenesis.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT peptide 1..21
FT /label= signal_peptide
FT Domain 26..76
FT /label= kunitz_domain_1
FT Misc-difference 36..40
FT /label= P1-P4'
FT /note= "motif replaced by Ile-Ala-Phe-Pro
FT to improve elastase inhibitory activity"

FT Domain 78..145
FT /label= kunitz_domain_2
FT Disulfide-bond 26..76
FT Disulfide-bond 35..59
FT Disulfide-bond 51..72
FT Disulfide-bond 82..132
FT Disulfide-bond 91..115
FT Disulfide-bond 107..128
FT Modified-site 10
FT /label= O-linked_glycosylation
FT Modified-site 45
FT /label= N-linked_glycosylation
XX

PN WO9603503-A1.
XX
PD 08-FEB-1996.
XX
XX
PF 21-JUL-1995; 95WO-JP01449.
XX
PR 21-JUL-1994; 94JP-0169221.
XX
PA (GREG) GREEN CROSS CORP.
XX
PI Goto T, Horii H, Ideno S;
XX
DR WPI: 1996-117048/12.
XX
PT Production of recombinant urinary trypsin inhibitor in Pichia sp. -
PT also kunitz domain fragments of the inhibitor and new variants
PT having improved elastase inhibitor activity

PS Disclousre; Fig 2; 97pp; Japanese.

XX
XX
CC Recombinant urinary trypsin inhibitor (rUTI) can be produced in
CC Pichia yeast. By mutating a 5 amino acid motif (P1-P4') within kunitz
CC domain 1 (from MGMTS to IAEFP), the resulting rUTI has improved
CC elastase inhibitory activity. The present sequence is that of
CC wild-type precursor UTI.

XX
SQ Sequence 147 AA:

Query Match 30.8%; Score 204; DB 17; Length 147;
Best Local Similarity 39.4%; Pred. No. 1.8e-13;
Matches 41; Conservative 12; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVLMHYDPQRCGCMTFPARCGDGAARGFTEYACQACARGPGDVCVPAVQ 67
DB 35 cmgmstsyf-----yngtmacetfgygcmgngnftvtekeclqctrtva--acnlpivr 88
OY 68 GRCRGHEPRMAYSPILQOCHPFYVGGCEGNGNMFHRSCECDAC 111
DB 89 gpcraflqlwaIdavkvcvllfpygqcgngnkfysekecreyc 132

RESULT 5
AAW25928
ID AAW25928 standard; protein; 147 AA.

XX
AC AAW25928;
XX
DT 11-NOV-1997 (first entry)
XX
DE Anti-trypsin inhibitor UTI.
XX
XX
KW Trypsin inhibitor; kunitz domain; protease; active site; elastase;
KW neurophil; disease; modification; site directed mutagenesis.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Modified-site 10
FT /label= O-linked glycosylation site
FT Domain 22..77
FT /note= "Kunitz domain 1"
FT Disulfide-bond 26..76
FT Disulfide-bond 35..59
FT Active-site 36
FT /note= "active site residue"
FT Disulfide-bond 51..72
FT Domain 78..147
FT /note= "Kunitz domain 2"
FT Disulfide-bond 82..132
FT Disulfide-bond 91..115
FT Active-site 92
FT /note= "active site residue"


```

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-451939/48.
XX DR N-PSDB; AAS22648.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Example 4; Page 642-643; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, hemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.
XX
SQ Sequence 352 AA;

Query Match 30.8%; Score 204; DB 22; Length 352;
Best Local Similarity 39.4%; Pred. No. 4.5e-13;
Matches 41; Conservative 12; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVLMHNDPQRGGCMTFPARCGDGAARGFETYACQACARCPGACVLPAYQ 67
DB 240 cmgmstrf---yngtsmacetfgygcmgngmfvtekeclqcrtya--acnlpivr 293

OY 68 GPCRGMERPMAYSPILQOCHPFVYGGCEGNGNMFHSRESCEDAC 111
DB 294 gpcrafiqlwaiddavkycvlfpryggcgngnkfysekreyc 337

RESULT 8
AAU14579
ID AAU14579 standard; Protein: 352 AA.
XX
AC AAU14579;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #450.
XX
KW Human; novel protein; Antianemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; hemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX

```

```

OS Homo sapiens.
XX
XX WO20015437-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-0502623.
XX
XX 25-JAN-2000; 2000US-0491404.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-451939/48.
XX
XX DR N-PSDB; AAS22684.
XX
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Example 4; Page 882; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, hemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.
XX
SQ Sequence 352 AA;

Query Match 30.8%; Score 204; DB 22; Length 352;
Best Local Similarity 39.4%; Pred. No. 4.5e-13;
Matches 41; Conservative 12; Mismatches 45; Indels 6; Gaps 2;

OY 8 CTGPTSPHLVLMHNDPQRGGCMTFPARCGDGAARGFETYACQACARCPGACVLPAYQ 67
DB 240 cmgmstrf---yngtsmacetfgygcmgngmfvtekeclqcrtya--acnlpivr 293

OY 68 GPCRGMERPMAYSPILQOCHPFVYGGCEGNGNMFHSRESCEDAC 111
DB 294 gpcrafiqlwaiddavkycvlfpryggcgngnkfysekreyc 337

RESULT 9
AAE05095
ID AAE05095 standard; Protein: 352 AA.
XX
AC AAE05095;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human Inter-alpha trypsin inhibitor (ITI) light chain.
XX

```

[illegible]

Db 240 cmgmstrf---yngtmaceltqygscmgngnfnftekeolqctctva--acnlpivr 293

QY 68 GPCRGMERPMAYSPLLQOCHPFYVGGCGEGRGNMNFHSRSECDAC 111
||||| 11: : : | |||||1111 : : : :
Db 294 gpcratfqlwaftavkvcvlfpygcqngnklfysekreycv 337

RESULT 10
AAB43730
ID AAB43730 standard; Protein: 366 AA.
XX
AC AAB43730;
XX
XX 08-FEB-2001 (first entry)
DE Human cancer associated protein sequence SEQ ID NO:1175.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
XX Homo sapiens.
XX OS
XX WC0200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MC-US05882.
XX
XX 12-MAR-1999; 9905-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX PI
XX WPI; 2000-587533/55.
XX DR N-PDB; AACT7939.
XX
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -
XX
XX Claim 11: Page 1797-1798; 2352pp; English.
XX
XX AACT7607 to AACT8448 encode the human cancer associated proteins given
XX in AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnerability; immunomodulator;
XX antidiabetic; antisthmatic; antirheumatic; antiarthritic;
XX antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;
XX dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
XX nootropic; vasotropic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AACT8449 to
XX AACT8457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 366 AA;
XX

CC This is the amino acid sequence of a novel protease inhibitor based on

FT	Misc-difference	55		wild type sequence"
FT	/note=	"mutated amino acid; replaces amino acid Gln in		
FT		wild type sequence"		
FT	Domain	78..147		
FT	/note=	"Kunitz domain 2"		
FT	Disulfide-bond	82..132		
FT	Disulfide-bond	91..115		
FT	Active-site	92		
FT	/note=	"active site residue"		
FT	Disulfide-bond	107..128		
PN	JP09124700-A.			
PD	13-MAY-1997.			
XX				
XX	07-NOV-1995;	95JP-0288527.		
PF				
PR	07-NOV-1995;	95JP-0288527.		
PA	(GREC) GREEN CROSS CORP.			
XX				
XX	WPI: 1997-316576/29.			
DR	N-PSDB: AAT79083.			
PT	New protease inhibitor - useful for treating diseases involving			
PT	elastase			
XX	Claim 8; Page 4; 37pp; Japanese.			
PS				
XX	This is the amino acid sequence of a protease inhibitor construct based			
CC	on the anti-trypsin inhibitor UTI (AAW25928). The novel inhibitor			
CC	contains the Kunitz domains 1 and 2 of UTI with several amino acid			
CC	substitutions: amino acids 22-25 of the wild type protein (KEDS) are			
CC	replaced by the sequence RPDF: the active site sequence and adjacent			
CC	amino acids (amino acids 36-40 of the wild type sequence) are replaced by			
CC	the sequence IAEFP; and amino acids 52 and 55 are also substituted. The			
CC	modified protease inhibitors are targeted to the protease elastase,			
CC	especially from neutrophils and can be used to treat diseases associated			
CC	with elastase. Modifications of the active site were done by site			
CC	directed mutagenesis.			
SO	Sequence 145 AA;			
	Query Match	30.3%; Score 201; DB 18; Length 145;		
	Best Local Similarity	38.2%; Pred. NO. 3.5e-13;		
	Matches 39; Conservative 12; Mismatches 49; Indels 2; Gaps 1;			
OY	10 GPTSPHLVMHYDPORGCGMPFAPRCDSGAARGFETYEACQACARGPADACVLPAVOGP 69			
DB	33 gpciaefryfyngtsmaccqltyvgscmgngnfnflekectlqcttva--acnlplavrgp 90			
OY	70 CRGWEPRAWYSFLLOOCHPEFYVGCEEGGNFNHRSBSCDAC 111			
DB	91 crafqlwaifdavgkcvlifpygscqgnknkysekcrecy 132			
	RESULT 15			
ID	AAW25937 standard; Protein: 165 AA.			
XX				
XX	AAW25937;			
XX				
DT	12-NOV-1997 (first entry)			
DE	Elastase inhibitor Epi-UTI-RPDF-52-55 from plasmid pHH35.			
XX				
KW	Trypsin inhibitor; Kunitz domain; protease; active site; elastase;			
XX	neutrophil; disease; modification; site directed mutagenesis.			
XS	Synthetic.			
XX				

FH	Key	Location/Qualifiers
FT	Peptide	1..20
FT	/note= "signal peptide"	
FT	Protein	21..165
FT	/note= "mature protein"	
FT	Modified-site	30
FT	/label= O-linked glycosylation site	
FT	Domain	42..97
FT	/note= "Kunitz domain 1"	
FT	Misc-difference	42..45
FT	/note=	"mutated amino acids: replaces wild type sequence KEDS"
FT	Disulfide-bond	46..96
FT	Disulfide-bond	55..79
FT	Misc-difference	56..60
FT	/note=	"mutated region; amino acids substituted for wild type amino acid sequence MGMTS; this includes the active site residue of domain 1"
FT	Disulfide-bond	71..92
FT	Misc-difference	72
FT	/note=	"mutated amino acid: replaces amino acid Glu In wild type sequence"
FT	Misc-difference	75
FT	/note=	"mutated amino acid: replaces amino acid Gln In wild type sequence"
FT	Domain	98..165
FT	/note= "Kunitz domain 2"	
FT	Disulfide-bond	102..152
FT	Disulfide-bond	111..135
FT	Active-site	112
FT	/note= "active site residue"	
FT	Disulfide-bond	127..148
PN	JP09124700-A.	
PD	13-MAY-1997.	
XX		
PF	07-NOV-1995;	95JP-0288527.
PR	07-NOV-1995;	95JP-0288527.
PA	(GREC) GREEN CROSS CORP.	
DR	WPI; 1997-316576/29.	
N-PSDB:	AAT79087.	
PT	New protease inhibitor - useful for treating diseases involving elastase	
PS	Disclosure: Fig 34; 37pp; Japanese.	
CC	This is the amino acid sequence of the polypeptide encoded by insert of plasmid pHH35 which comprises the novel elastase specific inhibitor Epi-Urt-RDPF-S2-95 (AAW53935). The inhibitor sequence is linked downstream of the yeast invertase (SUC2) signal peptide sequence. The modified protease inhibitors are targeted to the protease elastase, especially from neutrophils and can be used to treat diseases associated with elastase. Modifications of the active site were done by site directed mutagenesis.	
SQ	Sequence	165 AA:
Query Match	Best Local Similarity	30.3%; Score 201; DB 18; Length 165;
Matches	39; Conservative	12; Mismatches 49; Indels 2; Gaps 1.
OY	10 GPSTSHLVLMHPDPRGGCMTFPARGCDGAARGFETYEAQQACARGPADCVLPVAGCP	69
Db	cgsciaiffpyifngtscmaqcitfyvgcmgnnfmvtekecljctciva--acnlplavrgp	110
OY	70 CGMEPRMAYSPLLOOCHPEFYVGCEGCNGNNFNHSRESCDAC	111

Db 111 crafidawdawkvkvlfpygqgngnkfysekreyc 152

Search completed: February 26, 2002, 01:26:45
Job time: 319 sec

. . .

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:49:26 ; Search time 144.8 Seconds

(Without alignments)
521.247 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_548

Perfect score: 2855
Sequence: 1 CPNOLSPMLWMDAOSTCERE.....KKIIELEKQACELLNRPQD 516

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTRMBL17:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_proteob:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	294.5	10.3	3198	5	Q9U8G8	Q9U8G8 manduca sex
2	256	9.0	3060	5	Q9VAV4	Q9VAV4 drosophila
3	247.5	8.7	2167	5	Q76840	Q76840 caenorhabd
4	243.5	8.5	2174	5	Q9GQ80	Q9GQ80 drosophila
5	243	8.5	1572	5	Q44938	Q44938 haemochus
6	229	8.0	2225	5	Q45881	Q45881 caenorhabd
7	218	7.6	1043	5	Q17644	Q17644 caenorhabd
8	216.5	7.6	144	11	Q9QW87	Q9QW87 mesocricetu
9	213	7.5	151	4	P78491	P78491 homo sapien
10	211	7.4	922	5	Q21418	Q21418 caenorhabd
11	203	7.1	349	11	Q9DBJ9	Q9DBJ9 mus musculu
12	202.5	6.9	1743	5	Q9XWXS	Q9XWXS caenorhabd
13	196	6.9	352	11	Q70160	Q70160 cavia porce
14	195	6.8	1965	5	Q61893	Q61893 caenorhabd
15	194.5	6.8	287	13	Q93424	Q93424 cyprinus ca
16	191.5	6.7	1474	5	Q62504	Q62504 caenorhabd
17	189	6.6	2150	5	Q65428	Q65428 homo sapien
18	188.5	6.6	1280	11	Q94131	Q94131 caenorhabd
19	186	6.5	1280	11	Q9EPX2	Q9EPX2 mus musculu

20	185.5	6.5	4123	4	Q75851	Q75851 homo sapien
21	184.5	6.5	246	11	Q9Z208	Q9Z208 mus musculu
22	183.5	6.4	342	13	P70004	P70004 xenopus lae
23	182	6.4	187	5	P82968	P82968 mellithea c
24	179.5	6.3	1208	5	Q23456	Q23456 caenorhabd
25	177	6.2	763	5	Q9X2D0	Q9X2D0 drosophila
26	177	6.2	1539	5	Q09983	Q09983 caenorhabd
27	175.5	6.1	2026	4	Q00468	Q00468 homo sapien
28	175	6.1	396	6	Q28874	Q28874 canis famli
29	173.5	6.1	251	4	Q95103	Q95103 homo sapien
30	173	6.1	1549	11	Q60444	Q60444 cricetus
31	172	6.0	1344	11	Q9Z214	Q9Z214 mus musculu
32	171.5	6.0	947	5	Q44171	Q44171 caenorhabd
33	171.5	6.0	947	5	Q26262	Q26262 caenorhabd
34	170.5	6.0	838	5	Q27422	Q27422 caenorhabd
35	170.5	6.0	838	5	Q18761	Q18761 caenorhabd
36	170	6.0	67	11	Q9QW86	Q9QW86 mesocricetu
37	169.5	5.9	1419	13	Q98SW3	Q98SW3 brachydantio
38	169	5.9	1391	5	Q19021	Q19021 drosophila
39	168.5	5.9	1395	5	Q44924	Q44924 drosophila
40	165.5	5.8	1335	5	Q9W213	Q9W213 drosophila
41	164.5	5.8	1273	5	Q44928	Q44928 caenorhabd
42	162	5.7	423	5	P91572	P91572 caenorhabd
43	161.5	5.7	1297	5	Q9U350	Q9U350 caenorhabd
44	160	5.6	685	6	Q9TTS5	Q9TTS5 bos taurus
45	159.5	5.6	1380	4	Q9HCK4	Q9HCK4 homo sapien

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	3198 AA.
Q9U8G8	Q9U8G8			
AC	Q9U8G8			
DT	01-MAY-2000 (TREMREL. 13, Created)			
DT	01-MAY-2000 (TREMREL. 13, Last sequence update)			
DT	01-JUN-2001 (TREMREL. 17, Last annotation update)			
DE	LACUNIN PRECURSOR.			
OS	Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;			
OC	Sphingioidea; Sphingidae; Sphinginae; Manduca.			
OX	NCBI_TaxID=7130;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=9945716; PubMed=10528409;			
RA	Nardi J.B., Martos R., Maiden K.K., Lampe D.J., Robertson H.M.;			
RT	"Expression of lacunin, a large multidomain extracellular matrix			
RT	protein, accompanies morphogenesis of epithelial monolayers in Manduca			
RT	sexta.";			
RL	Insect Biochem. Mol. Biol. 29:883-897(1999).			
CC	-1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX			
CC	DOMAIN.			
CC	-1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.			
CC	EMBL; AF078161; AF04457.1; -.			
DR	HSSP; P12111; ZKNT.			
DR	InterPro; IPR003598; Ig_C2.			
DR	InterPro; IPR003006; Ig_MHC.			
DR	InterPro; IPR002223; Kunitz_BPTI.			
DR	InterPro; IPR000884; TSPI.			
DR	InterPro; IPR002221; WAP.			
DR	Pfam; PF00047; Ig; 2.			
DR	Pfam; PF00014; Kunitz_BPTI; 9.			
DR	Pfam; PF00095; wap; 1.			
DR	PRINTS; PR00759; BASICPTASE.			
DR	SMART; SM00408; IGC2; 2.			
DR	SMART; SM00331; KU; 10.			
DR	SMART; SM00209; TSPI; 7.			
DR	SMART; SM00217; WAP; 1.			
DR	PROSITE; PS00317; 4-DISULFIDE CORE; 1.			
DR	PROSITE; PS00280; BPTI_KUNITZ_1; 8.			
DR	PROSITE; PS00279; BPTI_KUNITZ_2; 10.			

DR PROSITE: PS50092; TSP1; 1.
 KW Serine protease inhibitor; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 SQ SEQUENCE 3198 AA; 349366 MW; ABA4CDA59C0D9134 CRC64;

Query Match 10.3%; Score 294.5; DB 5; Length 3198;
 Best Local Similarity 26.2%; Pred. No. 1.1e-14;
 Matches 113; Conservative 40; Mismatches 171; Indels 107; Gaps 19;

OY 3 NOLSPNLTWDAOSTCRESRQDDCAAKKCCINVCGLHCVAAAPPGSPAPPTTASAE 62
 DB 1709 NILSSIAVENCTSTFEFGCFDPTA-----TGPOEGGCPNSTERGCCP 1754
 OY 63 GFVCPDQSGDCDIDWGPAPCRDCEKPEPFTCASDGLTYNRR-----CYWDAEACLR 116
 DB 1755 DGVSPARGND---FEG-----CVITCNLS-STGCCPDGETPAHGPDDLCCLLSATGCCP 1805
 OY 117 -----GLHIVPCKHVLSMPSPSG--PEPTARPPGAAPVPALYSFSPQAVGV 167
 DB 1806 DNRKPAVGHLEGGCGQY-----SSFGCCPDNATVARGPN-----FQGGCGCYTEH 1851
 OY 168 GSTAHLHGVSGRP-----PPATYTWEX-QSHQRENILMRDQMTGNVYVS 212
 DB 1852 GCCPDHTEPAAPDYDGGCGCHTYEGCCPDGYTAKGNHGGDC---RDSQYG-----C 1903
 OY 213 IGLVLYNARPDAGLYTCTARNAA---GLIRA-----DFPLSVQREPARDAPS 260
 DB 1904 CGDGKTPATGPEREGDCATSEYGCPCDGLTEAKGHKEYGCGDLPEN-----KQAAAG 1956
 OY 261 IPAPAECLPDVQACTGPTSPHLYLMHYDQRGCMTFPARCDDGAARGETTEACQACA 320
 DB 1957 LPHD-----RGTCRNYSVYWYDYDEYGGCSFWMGGCEGNKKAETKECEDVCV 2006
 OY 321 R-GPDAVLPVAVOPRCRGMEBRMAYSPLLQOCHPVTYGGGNGNHNHRSRESCACGV 379
 DB 2007 QAPADACLPVAKGACLAGINIKMTYDABQEQCSQPIYGGCLGNANNTASLDQCORCCP 2066
 OY 380 PRTPPCRACRL 390
 DB 2067 ERSED--QCRL 2075

RESULT 2
 OYAV4 PRELIMINARY; PRT; 3060 AA.
 ID OYAV4
 AC OYAV4; OYAV4; OYAV4;
 DT 01-MAY-2000 (Tremblrel. 13. Created)
 DT 01-MAR-2001 (Tremblrel. 16. Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17. Last annotation update)
 DE CG1540 PROTEIN.
 GN PPN OR CG1540 OR CG18436.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-BEKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celaniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abbill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.F., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC -I- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL; AE003765; AAF56794.2; -;
 DR EMBL; AE003765; AAF56795.2; -;
 DR HSSP; P12111; 2KMT.
 DR FLYBASE; FBgn0003137; Ppn.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR003598; I9_C2.
 DR InterPro; IPR003006; I9_MHC.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR002221; WAP.
 DR Pfam; PF00047; I9; 3.
 DR Pfam; PF00014; Kunitz_BPTI; 12.
 DR Pfam; PF00090; tsp_1; 5.
 DR Pfam; PF00095; wap; 1.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00408; IGC2; 3.
 DR SMART; SM00131; KU; 12.
 DR SMART; SM00209; TSP1; 7.
 DR SMART; SM00217; WAP; 1.
 DR PROSITE; PS00317; 4.DISULFIDE CORE; 1.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 11.
 DR PROSITE; PS00279; BPTI_KUNITZ_2; 12.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS50092; TSP1; 3.
 KW Alternative splicing; Serine protease inhibitor.
 FT VARSPLIC 2803 2803 L->SVYVP (IN SHORT ISOFORM).
 FT VARSPLIC 2803 2803 ENFKTMSGVI -> VASRPPLHPNAV (IN SHORT
 FT VARSPLIC 2844 2854 ISOFORM).
 FT VARSPLIC 2855 3060 MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 3060 AA; 331579 MW; ACA31D3EE58C7C0 CRC64;

Query Match 9.0%; Score 256; DB 5; Length 3060;
 Best Local Similarity 20.6%; Pred. No. 1.2e-11;
 Matches 113; Conservative 41; Mismatches 203; Indels 192; Gaps 23;
 OY 1 CPNOSPNL-----WVDAOSTCERE-----CSRQDCAAEKCCINVC 38
 DB 1800 CPQYAKICELPAVGECAVTYSWYDTQDACRGTYGGCGGNGENKRPFEESCLARC 1859
 OY 39 GLHSCVAAPFGSPAPPTTASCEGFVCPDQSGDDCIW-----DGQVPCRCRDCEKPS 93

```

Db 1860 DRKEPTTTTPATRPQPSRODVCEBPAP---GECSTWLKMHFDRKIGACRQ----- 1909
QY 94 FTGASGLIYNNRC-----YMDAELACRLHLHIYPCKHVLSMPSPSPPEPTTARPT 147
Db 1910 -----EYFGNGCGNGNRFETENDCQ-----ORCLSOEPPATPPAPA-PTR 1950
QY 148 GAAPVPP-ALYSSPSP-----QAVOVGS-----TASLH 174
Db 1951 QPDPATYVACSQSPADPGQCDKALHMVNETGRCQSFYGGCGGNDNFATEECSAR 2010
QY 175 CDVS-----GRPPAVTWKQSHQREN--LIMRPDMYGNV-----VYTS 212
Db 2011 CSVNIDIRIGADP-----VEHDTSKCFLAEPGNCYNNVTWFWYNSAELCDEVEYTG 2063
QY 213 IQLVLYNARPEDA-----GLYTCTARNAGL----- 240
Db 2064 CGGNANNVATTEECQNECDNDQTTICALPYVGRCSDSLRRWYEDERSGECHEEFTGCRC 2123
QY 241 -RADFP-----LSVQREPARDAAPSIAPAECL--PDVQACTGPTSPHLVLMHYDPORG 292
Db 2124 NRRNFEVSQSDCLNFCIGEPV--VEPSAPITYSCAEPPEAGECDNRT-----AMFYDSENM 2177
QY 293 GCMTFPAAGCDGAARGFETYEACQACARPG--DACVLPVAVGQPCRGWEPMAYSPLLOQ 351
Db 2178 ACTAFTYTCGCGNGNRFETRDQCEKGVYDVCNEPVTGTCIDWQTKYFENTASQA 2237
QY 352 CHFTYVCGCEGNGNHNHRSCECDACPVPTPCRCRCLRSKALSLCSDFAIYGRLE 411
Db 2238 CEFTYVGGDGTGNRFSDSLFEQCTVCLAGREPR-----VGSARE 2276
QY 412 VLEEPEAG 420
Db 2277 ICLLPVATG 2285

```

```

RESULT 3
076840 PRELIMINARY; PRT: 2167 AA.
AC 076840: 022911;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE C37C3.6 PROTEIN.
GN C37C3.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitae; Pelodierinae; Caenorhabdilitis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gelsel C., Bredshaw H.;
RT "The sequence of C. elegans cosmid C37C3."
CC Submitted (JUL-1996) to the EMBL/GenBank/DDay databases.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC OF FORM B.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
CC EMBL: U64857; AAC25868.1; -.
CC EMBL: U64857; AAC25867.1; -.
CC HSSP: P00981; IPRK
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00047; Ig_1.
DR Pfam: PF00014; Kunitz_BPTI; 11.
DR Pfam: PF00090; TSP_1; 6.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00131; KU; 11.

```

```

DR SMART: SM00209; TSP1; 7.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 10.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 11.
DR PROSITE: PS00092; TSP1; 4.
FW Alternative splicing; Hypothetical protein; Serine protease inhibitor.
FT VARSPLIC 147 155 MISSING (IN ISOFORM A).
FT VARSPLIC 147 155 KDD -> SKF (IN ISOFORM A).
FT VARSPLIC 1556 1558 KDD -> SKF (IN ISOFORM A).
FT VARSPLIC 1559 2167 MISSING (IN ISOFORM A).
SQ SEQUENCE 2167 AA; 23759 MW; 96274786D52E3639 CRC64;

```

```

Query Match 8.7%; Score 247.5; DB 5; Length 2167;
Best local similarity 21.9%; Pred. No. 4,1e-11;
Matches 110; Conservative 45; Mismatches 167; Indels 181; Gaps 19;

```

```

QY 21 CSRDQCAAEEKCINVCGLHSCVAARFPSPAPPTTASCEGFVCPQDGSDDINDGDP 80
Db 1089 CNOTGSSG-----TYCGAGYKLAMHY-----DTGRCNQFWYGGCGNDNFFASQD 1135
QY 81 VC-----RC-----RDCEK--EPSTCASDGLTYNR-----CYM 109
Db 1136 MCETICEVPPGKRCYLPYVDGPLRCDQLOPRY-----YDHSKKHCVAFWMRGCLG 1187
QY 110 DA-----EACLGHLHLIYPCKHVLSW-----PPSSGPPEPTTAR--PTGAAPVPAL 156
Db 1188 NANNFNSFEFC-----SMFCKDVGPYDAPPTAAPPFPQQAQOYLPTPEVOLEIOS 1240
QY 157 YSSPSQAVO----- 166
Db 1241 AEQPOPQPPQOQOQOQOQOQPPQPSQMEDICRSRQDAGPCETYSQWFWYNAFSQECETFT 1300
QY 167 ---VGSTASL-----HCD-----VSGRPPPAVYWEK-----QSHQRNLLMRDQMY 205
Db 1301 YGCGGGLNLRNRSKDECEQCFVHGAQPSARQEAQPAQPAQPAQPSNYSPPQQA 1360
QY 206 GNVYVTSIGQLVLYNARPEDA-----GLYTCTARNAA 237
Db 1361 SPYVPS-----NSKQRPACHLNVDPQRCRKAQFDSWYEVATGSCVTPTKTYGCGGNAN 1413
QY 238 GLIRADFPISVQREPARDAAPSIAPAP--ECLPDVQACTGPTSPHLVLMHYDPORGCM 295
Db 1414 RFASKDQCESLCYKPASEAASAGIDGAAGINSYDEAKDQPCPTNFVTKYKKAQGTGN 1473
QY 296 TFPARGCDGAARGFETYEACQACARPGDACLVPVAVGQPCRGWEPMAYSPLLOCHFP 355
Db 1474 RFHYGCGQGTNNRFDNDQCKAAC--QNHKDACQLPVAVGQPCSKHSYTYTNSHOCETP 1532
QY 356 VYGCCEGNGNHNHRSCECDACP 378
Db 1533 TYGCGGLNTNRFATIECCQARCP 1555

```

```

RESULT 4
09G0R0 PRELIMINARY; PRT: 2174 AA.
ID 09G0R0:
AC 09G0R0:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EXTRACELLULAR MATRIX PROTEIN PAPILIN PRECURSOR.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DP CN BW;
RX PubMed-11076767;
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gulberg M., Kramer J.M., Achley B.D.,
RA Sieron A., Prockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the

```

RT ADAMTS metalloproteinases.
RL Development 127:5475-5485(2000).
CC -i SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL: AF205357; AAG37995.1; -
DR FLYBase; FBgn0003137; Ppn.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_1like.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSPL.
DR InterPro: IPR002221; WAP.
DR Pfam; PF00014; Kunitz_BPTI; 3.
DR PRINTS; PR00003; 4DISULPHCORE.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGc2; 2.
DR SMART; SM00410; IG_1like; 1.
DR SMART; SM00131; KU; 3.
DR SMART; SM00209; TSPL; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 3.
DR PROSITE; PS50092; TSPL; 3.
KW Matrix protein; Serine protease inhibitor; Signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 2174 AA; 231936 MW; 038E707952623120 CRC64;

Query Match 8.5%; Score 243.5; DB 5; Length 2174;
Best Local Similarity 24.7%; Pred. No. 8.5e-11;
Matches 109; Conservative 31; Mismatches 171; Indels 131; Gaps 23;

OY 1 CPNOLSPNLMWDAQ--SNCCERE--CSRDOCA--AEKCIWV--CGHSCVAARF 48
DB 1411 CPDQTPASGNGEGCCGCTERFGCCPDSTPAHGPKNKGGCLDTQCGCCDNLIAARG 1470
OY 49 PGSAPAP--TAASCEGFVCP-----QQGSDCDIDMGPVCRGRD----- 87
DB 1471 PNNEGCECHYTPSC---CPDNKSAATGYNQEGCACE--TQYGC-CPDKITAAKGRKH 1523
OY 88 ---CEKPRSTCASDGLTYNNRCYMDACLRGLHLHYVPRKH-----LSMPSPSP 136
DB 1524 EGCPE--TFQGCCPDGLTF-----AKGRHHGCHCTQTEFKCCDDEKTPAGRP 1571
OY 137 -GPETTRAPTPGAAP--VPALYSSPSPQAVOVGTASLHCDVSGRRP--PAVTEKOSH 192
DB 1572 NGEGETCHESKFGCCPDGVTKATDE-----KFGS---CENQEPKQKCGLPKEIG 1619
OY 193 QRENIMRP--DQYGNVVTSTIGQLVLYNARPD-----AGLYCTARNNAAGL 239
DB 1620 TCNNYSVYVYFDTSYGGCARFWYGGCDGNDNFSEAECKDTCODYTGKHYCLLPKRSAG- 1678
OY 240 LRADPPLSVQREPARDAAPSTIPAPAECLPDVQACTGPTSHLVLMHYDPQGGGMPRA 299
DB 1679 -----PCTGFTKK---WYFDVDNKRCEEFQY 1701
OY 300 RGCDAAGFETYEACQOACARGP--DACVLPVAVGPCRGWEPWAVSPLLQCHPFVYG 358
DB 1702 GGCYCTNNRPFDSLEQCGGTCASENLPCEQPVESGPCAGNERMYINDIDICRPTIYG 1761
OY 359 GCEGNGNNFHSRECEDACPV 380
DB 1762 GCKGNKNMYPTBHACNYNCRP 1783
RESULT 5
ID 044938 PRELIMINARY; PRT; 1572 AA.
AC 044938;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE THROMBOSPONDIN.
GN THRI.

OS Haemonchus contortus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida.
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCB1_Taxid=6289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOREDON;
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,
RA Knox D.P.;
RT "Cloning and characterization of thrombospondin, a novel multidomain
glycoprotein associated with the gut of Haemonchus contortus."
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -i SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AF043121; AAB99830.2; -
DR HSSP; P05067; ITAW.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSPL.
DR InterPro: IPR000884; TSPL.
DR Pfam; PF00014; Kunitz_BPTI; 6.
DR Pfam; PF00090; tsp_1; 6.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00131; KU; 6.
DR SMART; SM00209; TSPL; 7.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 6.
DR PROSITE; PS50092; TSPL; 2.
KW Serine protease inhibitor.
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2F903EC CRC64;

Query Match 8.5%; Score 243; DB 5; Length 1572;
Best Local Similarity 23.9%; Pred. No. 6.5e-11;
Matches 114; Conservative 45; Mismatches 175; Indels 142; Gaps 25;

OY 12 DAOSTCERECSDQCAAEKCCIVCGLS-----CVARF---PQ--SPAP 56
DB 770 ECVVTSSENGDVAAYVACSANATTTDLGLFNGTDTNATMCKSKSEFCPPWTFPAGRP 829
OY 57 TAASCEGFV---CPQGSQCDIWD-----GQPVCR-----CRDR----- 87
DB 830 NNAAGCPVYVLAGCNTEYEGCCCHDDVTLARGPVLBEGCGEPTCAGSLXGCKKDKTAFGPH 889
OY 88 ---CEKE-----PSFTCASDG-----LTYNNRCYMDAECRLGLHLHI 122
DB 890 YAGCERSSPPELSEFGCCADGETAALGANGTGCENCLTYKXGCCPGKGAAGHNEG 949
OY 123 VPCKHYVLSMPSPSPPEPTTAPTPGAAPVPALYSSPSPQAVOVGTASLHCDVSGRRP 182
DB 950 GCGVYA-QYGCCPDG--KTSAK---GA-----GFYGCPSCAQSOFG-----CCPDGKTP 993
OY 183 PAVTEKOSHORENLMRPDQ--MYG---NVVVTSTIGQLVLYNARPDAGL-----YT 230
DB 994 -----ARGSHKGS---CPQYTRYGCCPDGETTALG-----PRNDGCDCKYAYYG 1036
OY 231 CTARNAAGILRADPPLSVQREPARDAAPSTIPAPAECLPDVQACTGP-----TSPHLV 283
DB 1037 CCPDESKAIGPDYAGC-----PSTTLAPFLILGTVAPSKISSCALPDQGTGVSAGYKL 1091
OY 284 LMHYDPQGGGCTFPARGCDGAARGFETYEACQOACARGP--DACVLPVAVGPCRGWEP- 341
DB 1092 WYVYDTAAGRCQSQFWYGGCDGDNDFATKEQETTCVBPGLIGRGLRVKESPLCDDPQ 1151
OY 342 -RWAVSPLLQOCHPFVYVGGCEGNGNNFHSRECEDAC-----PVPRT 383
DB 1152 AKYWDYNTKQCAAFWVGCHGNMNNFASWECSTFFCQDVGFALPTAVAPQPP 1207

RESULT 6
ID 045881 PRELIMINARY; PRT; 2225 AA.
AC 045881;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)


```

QY 65 -----VCP00GSDCDIMDQVPCRCRCKEKESESTCASDGLTY 103
Db 697 FSYEGOKNANFLTEDEGLVCPVLPNPGSL--GEPILL--SLQEP-VICGGE--- 745
QY 104 YNRCYMDAECALRGHLHIVPCKHVLSPSSPSPETARTPGAAPVAPALYSPSPQ 163
Db 746 -----DTCNGNYCHV-----GGAPEET-NCCPGT-----RRPCDL 775
QY 164 AV0VG-GTASL-----HODVSGRPAPVATWEEKOSHORENLIMRPDOMYG 206
Db 776 PLEVQGVYKELERMFEDGIGMCRRFYVKGKMGNSNNFLT--KQSCROSKRMNCG-YG 832
QY 207 NVVYTSIGQLV-----YNAPEADAGLYCTARNAAGILRADPLSYVQRPAPDAAS 260
Db 833 DPLVYTERMLCTGGQRVNSCPQNS--YCHVGSSALFTLCC-----PKRKIDP- 879
QY 261 IPAPAECLPDVQACTGPRSPHLVLMHYDPQRGCMTPFARGCDGAARGETYEAOQA-- 318
Db 880 -----CDOAVEGTG--SEDLPRWFPRKONRCAPFYVGVAGNENNETISONTMEMACP 931
QY 319 -----CARGPDACVLPAYOGP 335
Db 932 EYRNYCPHIGILBESSTVTSQIDKCEPFGICMHSSEFNVSICCDPMDFCLASDQSP 991
QY 336 CRGMEPRNAYSPLLOQCHPEYVGGCEGNGNHFHRESCEDACVPRTPPCRACRLRSKU 394
Db 992 CNNEFKRYGDANDTQVCEYGGCEGTLLNHNHSLQRCETLC-----CKEYKRHRHL 1043

```

```

RESULT 8
ID 090M87 PRELIMINARY; PRT; 144 AA.
AC 090M87;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BIKUNIN.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN [1]
RP SEQUENCE.
RX MEDLINE=97018241; PubMed=8864857;
RA Yamamoto T., Yamamoto K., Sinohara H.;
RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian
RT hamster urine and plasma."
RL J. Biochem. 120:145-152(1996).
CC -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR HSSP; P02760; BIK.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI; 2.
DR PRINTS; PRO0759; BASICPTASE.
DR SMART; SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
KW Serine protease inhibitor.
SQ SEQUENCE 144 AA; 15700 MW; C71661C10FC2ACES CRC64;

```

```

Query Match 7.6%; Score 216.5; DB 11; Length 144;
Best Local Similarity 35.5%; Pred. No. 6.1e-10;
Matches 44; Conservative 16; Mismatches 57; Indels 7; Gaps 2;
QY 262 PAPAECPLPDVQAC-----TGPTSPHLVLMHYDPQRGCMTPFARGCDGAARGETYEAOQ 316
Db 14 PLYVTDVLAKKESQSLSYEGCLGIEKTYTNGASMACETIHYGGCLGNGNMFSEKCL 73
QY 317 QACARGPDACVLPAYOGPGRGMEPRNAYSPLLOQCHPEYVGGCEGNGNHFHRESCEDA 376
Db 74 QTCRTVA--ACSLPIVQPCRAVYELMAFDAAQKCVQFVYGGCGNGNKFYSECKEY 131

```

```

QY 377 CPVP 380
Db 132 CGVP 135

```

```

RESULT 9
ID P78491 PRELIMINARY; PRT; 151 AA.
AC P78491;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALPHA-1-MICROGLOBULIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91214554; PubMed=1708673;
RA Vetr H., Gebhard W.;
RT "Structure of the human alpha 1-microglobulin-bikunin gene."
RL Biol. Chem. Hoppe-Seyler 371:1185-1196(1990).
CC -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; X54818; CA38587.1; -.
DR EMBL; X54816; CA38587.1; JOINED.
DR HSSP; P02760; BIK.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI.
DR PRINTS; PRO0759; BASICPTASE.
DR SMART; SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
KW Serine protease inhibitor.
FT NON TER 1
SQ SEQUENCE 151 AA; 16542 MW; 88F400C5ECA19831 CRC64;

```

```

Query Match 7.5%; Score 213; DB 4; Length 151;
Best Local Similarity 40.2%; Pred. No. 1.2e-09;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;
QY 274 CTGPTSPHLVLMHYDPQRGCMTPFARGCDGAARGETYEAOQACARGPDACVLPAYO 333
Db 39 CMGMTSRFY-----YNGTSMACETIYGGCMGNGNMFYEKECLQTCRTVA--ACNLPIVR 92
QY 334 GPCRGMEPRNAYSPLLOQCHPEYVGGCEGNGNHFHRESCEDACVP 380
Db 93 GPCRAFIOLMAFDAYKGVLPYGGCGNGNKFYSECKEYCGVP 139

```

```

RESULT 10
ID Q21418 PRELIMINARY; PRT; 922 AA.
AC Q21418;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE K10D3.4 PROTEIN.
GN K10D3.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mcmurray A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

```


RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Garner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson M., Weinstock L., Wilkinson-Spoat J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: Z75545; CAA99886.1; -
 DR HSSP: P10646; ITEX.
 DR InterPro: IPR002899; EB.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF01683; EB; 1.
 DR Pfam: PF00014; Kunitz_BPTI; 5.
 DR PRINTS: PR00759; BASICPTASE.
 DR SMART: SM00131; KU; 5.
 DR SMART: SM00289; WRI; 6.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 5.
 KW Serine protease inhibitor.
 SQ SEQUENCE 922 AA; 9995 MW; 0CCBRC0AE2524CFD CRC64;

Query Match 7.4%; Score 211; DB 5; Length 922;
 Best Local Similarity 19.6%; Pred. No. 1,3e-08;
 Matches 108; Conservative 45; Mismatches 165; Indels 234; Gaps 24;

QY 1 CP-NQSLPLMND-----AQTCEKRECSDDQCAAEKCCINVCGLHSCVAARFPG 50
 DB 111 CPANGNNPFRIENOVQOCYGGADATCEAIGALSYDC-----ICSDSDCTVNN-PI 160
 QY 51 SPAAPTTASCS-----EGFVCPQGGSDCQIMDQPY-----CR-----CRDR 87
 DB 161 SFCCEPRACACIOPPNEGTPPEGGTTLNWHIDPITGCEKREKLYGCGNANNFOTKDH 220
 QY 88 CEKPEPFTCASDGLTYNN-----RCYMDAEACILGLHLITVPCKHYLSWPPSSP 136
 DB 221 CESYCKQTC-NRGLPLRYDRRTGVKQEPYCGGNDNGC-----NNP 260
 QY 137 GPETTARTPGAP-----VPPALYSSPSPQAVOVGTSALHCDV-----SGR 180
 DB 261 NVQCTTMTGLQCCPTLYLFCISRNNGIPEVYNT-----AGLPTFFDVGIPDGSGLN 313
 QY 181 PPAAVTWEEKSHQRENLIMRPDQMYGNVVTSTIGLVLYNAREDAGLYTCTARNAGLL 240
 DB 314 TSPRFYIDRE-----GRCIOPS-----YLGOGGNFNNFL 343
 QY 241 RADFLPLSVQR-----EPARDA-----APSIAPAPAC----- 267
 DB 344 SDDHCKFCFSRLCSAGEPLKDSGGERNMDCSPTSGANSCTPSCSTSGSTTFGVYC 403
 QY 268 -----LPVOACTGTSPHLVLMHYDPQNGGCMTPPARCCDAAGFETYEACQQA 318
 DB 404 CRRPOYVCKLPRBOQNCGYTNNR--WMFNAKTGNCSEEFYSGCQGNANNFETYYKKCCQDY 460
 QY 319 C--ARG-----PGDA--CY 328
 DB 461 CDDASEPCCIGTALTDSNGNFIICGGSASASTTPRANHYYCYDDTGTGCCPTQAYTCS 520
 QY 329 LPAVQG-PCRGWEPRAWYSPLLQOCHPFYVYGGCEGNGNFFHRSCEDEACPVPRTP----- 383
 DB 521 LSYKSGASGCPAVTRMYVYSTTRTCQTYSPFNGCDGNSNNFATQDDCKDYCRVESCSPDGE 580
 QY 384 -----PCRAC 388
 DB 581 VVKEDNGAARAC 592
 RESULT 11

Q9DBJ9
 ID Q9DBJ9 PRELIMINARY; PRT; 349 AA.
 AC Q9DBJ9;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE ALPHA 1 MICROGLOBULIN/BIKUNIN.
 GN AMBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawakawa T., Shinagawa A., Shidata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamane I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shidata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
 RA Wyshniuk-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RC Nature 409:685-690(2001).
 CC -1- SIMILARITY: TO PANCREATIC TRYPsin INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AK004907; BAB23659.1; -
 DR MGD: MGI:88002; Ambp.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR002345; Lipocalin.
 DR InterPro: IPR000566; Lipocalin_cytfabp.
 DR Pfam: PF00014; Kunitz_BPTI; 2.
 DR Pfam: PF00061; Lipocalin; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR PRINTS: PR00179; LIPOCALIN.
 DR SMART: SM00131; KU; 2.
 DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
 KW Serine protease inhibitor.
 SQ SEQUENCE 349 AA; 39101 MW; E593222FBC01BC3D CRC64;

Query Match 7.1%; Score 203; DB 11; Length 349;
 Best Local Similarity 40.6%; Pred. No. 2e-08;
 Matches 39; Conservative 13; Mismatches 42; Indels 2; Gaps 1;

QY 285 WHYDPQGGCMTPPARCDSGAAGFETYEACQACARGPDACVLPAVQGPCRGWEPRAW 344
 DB 246 YYNNGASMACETFOYGGCLGNGNFFISEKDCLOTCTRIA--ACNLPIVQGPCRAFLTMA 303
 QY 345 YSPLLQOCHPEFYVYGGCEGNGNFFHRSCEDEACPV 380
 DB 304 FDAAGCKCIOFHYGCGKNGNKFYSEKCEYCGVP 339
 RESULT 12
 Q9XMX5 PRELIMINARY; PRT; 1743 AA.
 AC Q9XMX5;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE Y43F8B.3 PROTEIN.
GN Y43F8B.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Ainscough R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE-94150718; PubMed=7906398;
RX Bonfield J., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins J., Kirsten T., Laister N., Latreille P.,
RA Jones M., Keshav J., Kirsten J., Laister N., Latreille P.,
RA Lighthouse J., Lloyd C., McMurtry A., Mottimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifen L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Therry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
RA Watson A., Weinstock L., Wilkison-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AL032623; CAA21511.1; -.
DR HSSP; P05067; ITAM.
DR InterPro; IPR002899; EB.
DR InterPro; IPR002223; Kunitz_BPTI.
DR Pfam; PF00014; Kunitz_BPTI; 16.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00131; KU; 15.
DR SMART; SM00289; WRI; 10.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 15.
KW Serine protease inhibitor
SQ SEQUENCE 1743 AA; 19111 MW; B93C163556433C2A CRC64;

Query Match 7.1%; Score 202.5; DB 5; Length 1743;
Best Local Similarity 22.1%; Pred. No. 1.3e-07;
Matches 124; Conservative 49; Mismatches 235; Indels 153; Gaps 26;
QY 20 ECSDQDQAAAEKCCINVCGLHSCVAARFPGSPAP-----TTAASCE 62
Db 1019 QGSPQGVCPGHHYCHVGTETANYCKA--LGDDPCGQPLDNGVGSQLSRWYWNQSQCL 1076
QY 63 GF-VCPQGG-----DDIDMDGQVYCRRCRCEKEPSTCASDGLTYNR---CYMD 110
Db 1077 PFSTCGGKGTONNEFLTKQDCD-----RTCYELDNPCALGDDPMAONNRPLOCSAT 1126
QY 111 AEACLRGLHILVPCKHLVSMPPSPGPPTTAPT---GGAAPVPALVSSPQAVOV 167
Db 1127 ASFGCAQFMCHEFGANDQFTVCCPGRVESFQICQPAVAGTGATLFWYTYNAOTMCQVOP 1186
QY 168 GGTASLHCDVSGRPRAVTEKOSHORENIMRPDMYGNVYVTSIGQLVLYNARPDAG 227
Db 1187 -----NYAGR-----MGNONNFQSOQACEQTCPIYVNVCPFGSPMLDASTNKP----- 1229
QY 228 LYTCT-ARNAGILRADF--PLSVQRE-PARDAAPSIAPACLEDPVQACTPTSPHLV 283
Db 1230 -VPTGFSNCG---ADHWCHLGLVDEYQCCPSPTNPACGLPDESEGVGAPAPPTS 1285
QY 284 LMHYDPOGCGMTFPAAGCGAARGFTYFACQOAC-----ARGPDA 326
Db 1286 RMYDDQDMCKQFTYNGRGNONNPLTODCACTDVFTNPNOPALPATLCSGGSS 1345
QY 327 -----CVLPVQGV---CRGWEPRMAYSPLLQOCHPEFVYGGCGNGNMFHRSRSCD 375
Db 1346 DTGAMWMCCHIGANDSTVCC-----PSGRSKSNFQOCLPQYQYNGKNGNMFENQADCR 1401

QY 376 ACVPFRP-----PCRACRLRSKLTALSLS--DFAIVGRLE----- 411
Db 1402 TCFVFINCLCEVILEDGAPKPC-----PLKNSCSASAEFCHTGPSPQSSFCCPR 1455
QY 412 VLEEP-----EAAGIARVALEDVLK-----DKMGLK-----FTGKYLEVTLGMD 454
Db 1456 INDPQNAFVANGSGNFMTRYNPNVPGDCFSQYRKLKGNENNFLTKMCQETCKPRT 1515
QY 455 WACPCPNMTAGDGLVIMGEV 475
Db 1516 TAC-----FGGESPLMNGRV 1531

RESULT 13
070160 PRELIMINARY; PRT; 352 AA.
ID 070160
AC 070160;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALPBA-1-MICROGLOBULIN/BIKUNIN.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.
OX NCBI_Taxid=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=LIVER.
RA Yoshida K., Suzuki Y., Yamamoto K., Shinohara H.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR EMBL; AB006444; BAA25305.1; -.
DR HSSP; P02760; IBIK.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR002343; Lipocalin.
DR InterPro; IPR000566; Lipocalin_cytfabp.
DR Pfam; PF00014; Kunitz_BPTI; 2.
DR Pfam; PF00061; Lipocalin; 1.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR00179; LIPOCALIN.
DR SMART; SM00131; KU; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Serine protease inhibitor.
SQ SEQUENCE 352 AA; 39033 MW; BAA7447B574C2FAB CRC64;

Query Match 6.9%; Score 196; DB 11; Length 352;
Best Local Similarity 36.2%; Pred. No. 7.2e-08;
Matches 38; Conservative 13; Mismatches 52; Indels 2; Gaps 1;
QY 276 GPTSPHLVLMHYDPQGGCGMTFPAAGCGAARGFTYFACQOACARGPDACVLPVQGP 335
Db 238 GPLGLEFRKRYFGSSMADEIFHYGCLGNNGNMFNSEKICQLQCRVYA--ACNLPIVGP 295
QY 336 CRGWEPRMAYSPLLQOCHPEFVYGGCGNGNMFHRSRSCDACPVP 380
Db 296 CQSSAQLMAFADAIKGCYVFTYGGCGNGKRYSEKEKEICGAP 340

RESULT 14
061893 PRELIMINARY; PRT; 1965 AA.
ID 061893
AC 061893;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE T07H8.4 PROTEIN.
GN T07H8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

XX Rhabdilitidae: Peiorderitinae: Ctenorhabdilitis.
 OC NCBI_TextID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copestey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hiller L., Jler M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latteille P.,
 RA Juthling J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Welshons L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Galtung S., Maggi L.;
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 CC -1 - SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL: AF067945; AAC17681.1; -.
 DR HSSP: P00981; 1DTK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI; 11.
 DR PRINTS: PRO0759; BASICPTASE.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00131; KU; 14.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 5.
 DR PROSITE: PS0279; BPTI_KUNITZ_2; 14.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_2.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR Calcium binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat;
 KW Serine protease inhibitor.
 WQ SEQUENCE 1965 AA: 223504 MW: 80E22ACB1C5E2234 CRC64;
 WQ

	Query Match	6.8%	Score 195;	DB 5;	Length 165;	
	Best Local Similarity	32.8%	Pred. No.	5.9e-07;		
	Matches	43;	Conservative	23;	Mismatches	61;
					Indels	4;
					Gaps	3
OY	284 LHMHPORGCGCTEPFARGCDGAARGETYEACQACARGDGYLPAVQGP-CRGWEPR	342				
Db	1748 VAFYMLTRGTCDFLYGGCGGNRFEDTFLIOKACEYGTGPCMESLDRGSWCAMSNR	1807				
OY	343 WAFSPLLQCHDFYVGCGEGNNHRSCECADP--VPRTTPC-RACRLSKLTALSLC	399				
Db	1808 YFFNKRAROCKCFHTTGCKSCSNNFLTKRECOTKEKKRRPLAAPSKKAKKLAKIITFEQ	1867				
OY	400 RSDFAIVGRLT	410				
Db	1868 NSRYNVVTGRDT	1878				
RESULT	15					
ID	093424					
AC	093424:	PRELIMINARY:	PRT:	287	AA.	
DT	01-NOV-1998 (TREMBLrel. 08. Created)					
DT	01-NOV-1998 (TREMBLrel. 08. Last sequence update)					

DT 01-JUN-2001 (TRENBLER, J), Last annotation update)
 DE HYPOTHEETICAL 33.1 KDA PROTEIN.
 OS Cyprinus carpio (Common carp).
 OC Euxaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
 OX NCBI_TaxId=7962;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Gracey A. Y.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
 DR EMBL; AF008648; AAC19410.1; -.
 DR HSSP; P31713; 1SHP.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR Pfam; PF00014; Kunitz_BPTI; 3.
 DR PRINTS; PR00759; BASICPTASE.
 DR SMART; SM00131; KU; 3.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 3.
 KW Hypothetical protein; Serine protease inhibitor.
 SQ SEQUENCE 287 AA; 33093 MW; DF693BD76718115E CRC64;

	Query Match	6.8%	Score 194.5	-DB 13	Length 287
	Best Local Similarity	35.9%	Pred. No. 7.6e-08		
	Matches 37	Conservative 14	Mismatches 51	Indels 1	Gaps 1
QY	276 GPTSHLVLMHNDPQRGGCMTFPARGCCDAAGFEIYEACQAC -ARGSGDACVLPAYOG 334				
	: : : : : : : : :				
Db	49 GKCKALKRFYEDTDTRGSEFSEYGCCQENNNFFETLQCEKCLVKEDSKSPQLDDEPG 108				
QY	335 PCRGMEPRMAYSPLLQQCHPFYUUGCGEGGNPNHRESCEDAC 377				
	: : : : : : : :				
Db	109 PGRGLVPRIYFDFKSQECKRFFYGGGFGGANNEFKTKKEHCRC 151				

Search completed: February 26, 2002, 01:49:37
Job time: 1286 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:43 : Search time 45.78 Seconds

(without alignments)
413.260 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_548
Perfect score: 2855
Sequence: 1 CPNQLSPNLMTWDAQSTCERE.....KKILELLEKQACELLNRFDD 516

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Swissprot_39:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	7.7	346	1 AMBP_MERUN	Q62577 meriones un
2	218.5	7.7	349	1 AMBP_MESAU	Q60555 mesocricetu
3	214	7.5	337	1 AMBP_PIG	P04366 sus scrofa
4	214	7.5	349	1 AMBP_RAT	Q64240 rattus norv
5	213	7.5	352	1 AMBP_HUMAN	P02760 homo sapien
6	210	7.4	352	1 AMBP_BOVIN	P00378 bos taurus
7	208	7.3	123	1 IATR_HORSE	P04365 equus cabal
8	205	7.2	123	1 IATR_SHEEP	P13371 ovis aries
9	204	7.1	300	1 TFPI_RABIT	P19761 oryctolagus
10	203	7.1	349	1 AMBP_MOUSE	Q07456 mus musculu
11	194.5	6.8	304	1 TFPI_HUMAN	P10646 homo sapien
12	192	6.7	252	1 SPPI2_HUMAN	O43391 homo sapien
13	192	6.7	252	1 SPPI2_MOUSE	O9WU03 mus musculu
14	186.5	6.5	306	1 TFPI_MOUSE	O54819 mus musculu
15	184.5	6.5	302	1 TFPI_RAT	Q02445 rattus norv
16	181.5	6.4	1416	1 YNBI_CAEEL	Q03610 caenorhabdi
17	179.5	6.3	235	1 TFPI_HUMAN	P48307 homo sapien
18	179.5	6.3	304	1 TFPI_MACMU	Q28864 macaca mula
19	176.5	6.2	2944	1 CA17_HUMAN	Q02388 homo sapien
20	176	6.2	110	1 IBP_CARCR	P00993 carecta car
21	176	6.2	1559	1 AGRI_RAT	P25054 rattus norv
22	173.5	6.1	513	1 SPPI1_HUMAN	O43378 homo sapien
23	167.5	5.9	62	1 IP33_ANESU	P10380 anemonda su
24	166.5	5.8	353	1 AMBP_PLEPL	P36992 pleuronecte
25	164	5.7	1955	1 AGRI_CHICK	P31696 gallus gall
26	161	5.6	230	1 TFPI2_MOUSE	O35536 mus musculu
27	158	5.5	61	1 IBPI_TACTR	P16044 tachyples
28	158	5.5	507	1 SPPI1_MOUSE	O91097 mus musculu
29	158	5.5	1051	1 PRK7_CHICK	O91048 gallus gall
30	158	5.5	1070	1 PRK7_HUMAN	O13308 homo sapien
31	157.5	5.5	2481	1 UN52_CAEEL	Q06561 caenorhabdi
32	157.5	5.5	3176	1 CA36_HUMAN	P12111 homo sapien
33	155.5	5.4	4393	1 PGBM_HUMAN	P98160 homo sapien

ALIGNMENTS

```

RESULT 1
ID      AMBP_MERUN      STANDARD:      PRT;      346 AA.
AC      062577; 062576;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      AMBP_PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE      TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
OS      AMBP OR ITIL.
GN      Meriones unguiculatus (Mongolian jird).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC      Meriones.
OX      NCBI_TaxID=10047;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RX      MEDLINE=95110820; PubMed=7529051;
RA      Ide H., Itoh H., Nawa Y.;
RT      Sequencing of cDNAs encoding alpha 1-microglobulin/bikunin of
RT      Mongolian gerbil and Syrian golden hamster in comparison with man and
RT      other species.
RL      Blochim. Biophys. Acta 1209:286-292(1994).
CC      -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC      FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC      APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC      AND ALBUMIN (BY SIMILARITY).
CC      -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
CC      URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOZOMAL GRANULOCYTIC
CC      ELASTASE (BY SIMILARITY).
CC      -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC      SEPARATELY FUNCTIONING PROTEINS.
CC      -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
CC      SIMILARITY).
CC      -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN
CC      FAMILY.
CC      -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sb.ch/announce/
or send an email to license@isb-sb.ch).
-----
EMBL: D31813; BA0600.1; -
HSSP: P10646; IAD6.
InterPro: IPR002223; Kunitz_BPTI.
InterPro: IPR002345; Lipocalin.
InterPro: IPR000566; Lipocalin_cyFABP.
Pfam: PF00014; Kunitz_BPTI; 2.
Pfam: PF00061; Lipocalin; 1.
PRINTS: PR00179; LIPOCALIN.
PRINTS: PR00759; BASICPTASE.

```

DR SMART: SM00131; KU: 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
Lipocalin.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 205 202 ALPHA-1 MICROGLOBULIN.
FT CHAIN 205 346 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
CHAIN.
FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
FT BINDING 286 336 BPTI/KUNITZ INHIBITOR 2.
FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
FT DISULFID 90 187 BY SIMILARITY.
FT DISULFID 230 280 BY SIMILARITY.
FT DISULFID 239 263 BY SIMILARITY.
FT DISULFID 255 276 BY SIMILARITY.
FT DISULFID 286 336 BY SIMILARITY.
FT DISULFID 295 319 BY SIMILARITY.
FT DISULFID 311 332 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
FT ACT_SITE 240 241 ELASTASE) (BY SIMILARITY).
FT ACT_SITE 296 297 INHIBITORY SITE (P1) (TRYPsin) (BY
SIMILARITY).
SQ SEQUENCE 346 AA; 38643 MW; F1A4463810918D5F CRC64;

Query Match 7.7%; Score 219; DB 1; Length 346;
Best Local Similarity 40.0%; Pred. No. 3.1e-07;
Matches 42; Conservative 13; Mismatches 48; Indels 2; Gaps 1;

OY 276 GPISPHLVLMHYDPORGCMTPPARGCDGAAAGETYEACQACARGDACYLPAVQGP 335
DB 237 GPCLGMERHYNGTSMACETFGYGGCLGNGNNFISEKECLOTCTVA--ACNLPYVQGP 294
OY 336 CRGMEPRMAYSPILQOCCHPFYGGCEGNGNMFHRESCEDACPV 380
DB 295 CRAVYKIMAFDAQCKCIQFTYGGCKGNGKPFSEKECKEYCGVP 339

RESULT 2
AMBP_MESAU STANDARD; PRT; 349 AA.
ID AMBP_MESAU STANDARD; PRT; 349 AA.
AC Q60559; Q60558; Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].
GN AMBP OR ITIL.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Liver;
RC MEDLINE=95110820; PubMed=7529051;
RA Ige H., Itoh H., Nawa Y.;
RT "sequencing of cDNAs encoding alpha 1-microglobulin/bikunin of
RT Mongolian gerbil and Syrian golden hamster in comparison with man and
RT other species.";
RL Biochim. Biophys. Acta 1209:286-292(1994).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN (BY SIMILARITY).
CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
CC ELASTASE (BY SIMILARITY).
CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO

CC SEPARATELY FUNCTIONING PROTEINS.
CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
CC SIMILARITY).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D31814; BAA06601.1; .
CC HSSP; P10646; 1TFX.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000566; Lipocalin_cyFPABP.
DR Pfam; PF00014; Kunitz_BPTI; 2.
DR Pfam; PF00061; Lipocalin; 1.
DR PRINTS; PR00179; LIPOCALIN.
DR PRINTS; PR00759; BASICPTASE.
DR SMART: SM00131; KU: 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00213; LIPOCALIN; 1.
KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
Lipocalin.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
FT CHAIN 205 349 INTER-ALPHA-TRYPsin INHIBITOR LIGHT
CHAIN.
FT DOMAIN 230 280 BPTI/KUNITZ INHIBITOR 1.
FT BINDING 286 336 BPTI/KUNITZ INHIBITOR 2.
FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
FT DISULFID 90 187 BY SIMILARITY.
FT DISULFID 230 280 BY SIMILARITY.
FT DISULFID 239 263 BY SIMILARITY.
FT DISULFID 255 276 BY SIMILARITY.
FT DISULFID 286 336 BY SIMILARITY.
FT DISULFID 295 319 BY SIMILARITY.
FT DISULFID 311 332 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
FT ACT_SITE 240 241 ELASTASE) (BY SIMILARITY).
FT ACT_SITE 296 297 INHIBITORY SITE (P1) (TRYPsin) (BY
SIMILARITY).
SQ SEQUENCE 349 AA; 38782 MW; 8C954584B7DBE728 CRC64;

Query Match 7.7%; Score 218.5; DB 1; Length 349;
Best Local Similarity 32.9%; Pred. No. 3.3e-07;
Matches 50; Conservative 24; Mismatches 67; Indels 11; Gaps 4;

OY 234 RNAAGLIRADPPLSVYQREPARDAAPSPAECLIPYQAC-----TGPTSPHLVLMHYD 288
DB 194 QKSTSLRA--RRAVLQGE--NEGSGTGPLVTDLKEDSCQLSYSGPCLGMTEKYYN 249
OY 289 PORGCMTPPARGCDGAAAGETYEACQACARGDACYLPAVQGPARGMEPRMAYSP 348
DB 250 GASMACTEFHYGGCLGNGNNSERKELOLCRTVA--ACSLPIVQGPCRAVVELMARDAA 307
OY 349 LQOCHPRVYGGCEGNGNMFHRESCEDACPV 380
DB 308 QGKCVQPSYGGCKGNGKFTSEKECKEYCGVP 339

RESULT 3
AMBP_PIG STANDARD; PRT; 337 AA.
ID AMBP_PIG STANDARD; PRT; 337 AA.

AC P04366; P34954;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMB PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKININ) (HI-30) (EI-14)]
 DE (FRAGMENT).
 GN AMB OR ITIL.
 OS Sus scrofa (pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_Taxid=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90353595; PubMed=1696914;
 RA Gebhard W., Schreimüller T., Vetter H., Wachter E., Hochstrasser K.;
 RT "Complementary DNA and deduced amino acid sequences of porcine alpha
 RT 1-microglobulin and bikunin.";
 RL FEBS Lett. 269:32-36(1990).
 RN [2]
 RP SEQUENCE OF 2-337 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=91113729; PubMed=1703444;
 RA Tavaekol A.;
 RT "Molecular cloning of porcine alpha 1-microglobulin/HI-30 reveals
 RT developmental and tissue-specific expression of two variant messenger
 RT ribonucleic acids.";
 RL Biochim. Biophys. Acta 1088:47-56(1991).
 RN [3]
 RP SEQUENCE OF 212-334.
 RA MEDLINE=85225967; PubMed=2408637;
 RT Huchtrassner K., Wachter E., Albrecht G.J., Reisinger P.;
 RT "Kunitz-type protease inhibitors derived by limited proteolysis of
 RT the inter-alpha-trypsin inhibitor, X. The amino-acid sequences of the
 RT trypsin-released inhibitors from horse and pig inter-alpha-trypsin
 RT inhibitors.";
 RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN.
 CC -1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE.
 CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X53685; CA37725.1; -;
 DR EMBL; X52087; CA36306.1; -;
 DR PIR; A01208; TIRPBI.
 DR PIR; S11056; S11056.
 DR HSSP; P10646; ITPX.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR000566; Lipocalin_cytfabp.
 DR Pfam; PF00014; Kunitz_BPTI; 2.
 DR Pfam; PF00061; Lipocalin; 1.
 DR SMART; SM00131; KU; 2.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE; PS0279; BPTI_KUNITZ_2; 2.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;

KW Lipocalin.
 FT NON_TER 1 1
 FT SIGNAL <1 188
 FT CHAIN 5
 FT CHAIN 191 337
 FT DOMAIN 216 266
 FT DOMAIN 272 322
 FT BINDING 38 38
 FT BINDING 76 173
 FT DISULFID 216 266
 FT DISULFID 225 249
 FT DISULFID 241 262
 FT DISULFID 272 322
 FT DISULFID 281 305
 FT DISULFID 297 318
 FT DISULFID 297 318
 FT CARBOHYD 100 100
 FT CARBOHYD 235 235
 FT ACT_SITE 226 227
 FT ACT_SITE 282 283
 FT ACT_SITE 49 49
 FT CONFLICT 259 259
 FT CONFLICT 270 270
 FT CONFLICT 278 278
 FT CONFLICT 283 283
 FT CONFLICT 285 286
 FT CONFLICT 293 293
 FT CONFLICT 311 311
 FT CONFLICT 315 315
 FT SEQUENCE 337 AA; 37650 MW; 1F630F983CD70F CRC64;
 Query Match 7.5%; Score 214; DB 1; Length 337;
 Best Local Similarity 39.0%; Pred. No. 6.1e-07;
 Matches 41; Conservative 15; Mismatches 47; Indels 2; Gaps 1;
 QY 276 GPTSPHLYMHNDPQGGCTPARGCDGANGFEYKACQACANGPDACYLPVQGP 335
 DB 223 GGLGMIKRYFYNGSSMADETFHYGGCMGNFNVESEKCLPTCR--TVEACSLPIVSGP 280
 QY 336 CGWEPFRWYSPDLOOCHPEFYVGGCEGNGNPNFRESCEDCACPVP 380
 DB 281 CRGFQPLMAFDVAGKCYLFENYGGCGNGNGNYSEKCEYECYCV 325
 RESULT 4
 AMB_RAT STANDARD; PRT; 349 AA.
 AC 064240; 063336;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMB PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
 DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKININ) (HI-30)].
 GN AMB OR ITIL.
 OS Rattus norvegicus (rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92182014; PubMed=1371936;
 RA Lindqvist A., Bratt T., Altiert M., Kasterm W., Aakerstrom B.;
 RT "Rat alpha 1-microglobulin: co-expression in liver with the light
 RT chain of inter-alpha-trypsin inhibitor.";
 RL Biochim. Biophys. Acta 1130:63-67(1992).
 RN [2]
 RP SEQUENCE OF 141-195 FROM N.A.
 RA MEDLINE=87033744; PubMed=2429963;
 RA Kasterm W., Björck L., Aakerstrom B.;
 RT "Developmental and tissue-specific expression of alpha 1-microglobulin

RT	MRNA	lnthe rat.";
RL	J. Biol Chem.	261.15070-15074(1986).
CC	-I-	FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC	CC	FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC	CC	APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC	CC	AND ALBUMIN (BY SIMILARITY).
CC	-I-	FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
CC	CC	URINE. INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOcytic
CC	CC	ELASTASE (BY SIMILARITY).
CC	-I-	P1M: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC	CC	SEPARATELY FUNCTIONING PROTEINS.
CC	-I-	P1M: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
CC	CC	SIMILARITY).
CC	-I-	SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC	CC	FAMILY.
CC	-I-	SIMILARITY: CONTAINS 2 BP1/KUNITZ INHIBITOR DOMAINS.
CC	CC	-----
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation - -
CC	CC	the European Bioinformatics Institute. There are no restrictions on its
CC	CC	use by non-profit institutions as long as its content is in no way
CC	CC	modified and this statement is not removed. Usage by and for commercial
CC	CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	CC	or send an email to license@isb-sib.ch).
CC	CC	-----
DR	EMBL:	S87544; AAB21782.1; -.
DR	EMBL:	J02600; AAA41596.1; -.
DR	HSSP:	P12111; IKUN.
DR	InterPro:	IPR002223; Kunitz_BP1.
DR	InterPro:	IPR002345; Lipocalin.
DR	InterPro:	IPR000566; Lipocaln_cytfabp.
DR	Pfam:	PF00014; Kunitz_BP1_2.
DR	Pfam:	PF00061; lipocalin; 1.
DR	PRINTS:	PR00179; LIPOCALIN.
DR	PRINTS:	PR00759; BASICPTASE.
DR	SMART:	SMO0131; KU; 2.
DR	PROSITE:	PS00280; BP1_KUNITZ_1; 2.
DR	PROSITE:	PS50279; BP1_KUNITZ_2; 2.
DR	PROSITE:	PS00213; LIPOCALIN; 1.
KW	Glycoprotein:	Plasma; Signal; Serine protease inhibitor; Repeat;
KW	Lipocalin.	
FT	SIGNAL	1 19 BY SIMILARITY.
FT	CHAIN	20 202 ALPHA-1-MICROGLOBULIN,
FT	CHAIN	205 349 INTER-ALPHA-TRYPSIN INHIBITOR LIGHT
FT	DOMAIN	230 280 CHAIN.
FT	BINDING	286 336 BP1/KUNITZ INHIBITOR 1.
FT	DISEUFD	52 52 BP1/KUNITZ INHIBITOR 2.
FT	DISEUFD	90 187 CHROMOPHORE (BY SIMILARITY).
FT	DISEUFD	230 280 BY SIMILARITY.
FT	DISEUFD	239 263 BY SIMILARITY.
FT	DISEUFD	255 276 BY SIMILARITY.
FT	DISEUFD	286 336 BY SIMILARITY.
FT	DISEUFD	295 319 BY SIMILARITY.
FT	DISEUFD	311 332 BY SIMILARITY.
FT	CARBOHYD	114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	ACT_SITE	240 241 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
FT	ACT_SITE	241 ELASTASE) (BY SIMILARITY).
FT	ACT_SITE	296 297 INHIBITORY SITE (P1) (TRYPSIN) (BY
FT	CONFLICT	142 142 G->A (IN REF. 2).
SO	SEQUENCE	349 AA; 38851 MW; 1B7FB/DCE0824E01 CRC64;

Query Match 7.5%; Score 214; DB 1; Length 349;
Best Local Similarity 40.6%; Pred. No. 6.4e-07;
Matches 39; Conservative 14; Mismatches 41; Indels 2; Gaps 1;

Qy 285 WHYDPQGGCMTFPAKCGDGAARGFETYTEACQQAACARGPDACVLPAYQGPCRGWEPRWA 344

Db 246 YYYNASMACCTFYQYSGCLGNGNNFASAEKELQYCRITIA--ACNLPYVGQPCRAFAELMA 303

```

QY      345 YSPLLQCHPEFVYGGCEGNGNHFHSRECEADACVP 380
      :  :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      304 FDAAGGCTQFIYGGCKGNKGFYSRECKEYCGVP 339

```

RESULT

ID	AMBP_HUMAN	STANDARD;	PRT;	352 AA
AC	P02760; P02759; P00977;			

DT	21-JUL-1986	(Rel. 01, Created)
DT	13-AUG-1987	(Rel. 05, Last sequence update)
DT	20-AUG-2001	(Rel. 40, Last annotation update)

DE AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN (PROTEIN HC)
DE (COMPLEX-FORMING GLYCOPROTEIN HETEROGENEOUS IN CHARGE); INTER-ALPHA
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)].

GN AMBP OR ITIL OR HCP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=91214554; PubMed=1708673;
RA Vetr H., Gebhard W.;

RT "Structure of the human alpha 1-microglobulin-bikunin gene." ;
 RL Biol. Chem. Hoppe-Seyler 371:1185-1196(1990).
 RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87040757; PubMed=2430261;

RA Kaumeyer J.F., Polazzi J.O., Kotlick M.P.;
RT "The mRNA for a proteinase inhibitor related to the HI-30 domain of
RT inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin

RT (protein HC).";
RL Nucleic Acids Res. 14:7839-7850(1986)
RN [3]

RP	SEQUENCE FROM N.A.
RC	TISSUE=Liver;
RX	MEDLINE=90336621; PubMed=1696200;

RA Diarra-Mehrpour M., Bourguignon J., Seboue R., Salter J.P.,
RA Leveillard T., Martin J.P.;
RT "Structural analysis of the human inter-alpha-trypsin inhibitor

RT light-chain gene,";
RL Eur. J. Biochem. 191:131-139(1990).
RN [4]

RP SEQUENCE OF 1-220 FROM N.A.
RX MEDLINE=86312901; Pubmed=2428011;
RA Traboni C., Cortese R.;

RT "Sequence of a full length cDNA coding for human protein HC (alpha 1
RT microglobulin).";
RL Nucleic Acids Res. 14:6340-6340(1986).

RN [5]
RP SEQUENCE OF 20-202 (INDIVIDUAL WITH TUBULAR PROTEINURIA)
RX MEDLINE=84126849; PubMed=6198962;

RT The complete amino acid sequence of human complex-forming glycoprotein heterogeneous in charge (protein HC) from one

RT individual.";
RL Arch. Biochem. Biophys. 228:544-554 (1984)
RN [6]

RP SEQUENCE OF 20-198 (VARIANT).
RA Lopez C., Grubb A.O., Mendez E.;
RT "Human protein HC displays variability in its carboxyl-terminal amino

RT	acid sequence,"
RL	FEBS Lett. 144:349-353(1982).
RN	[7]

RP SEQUENCE OF 20-198 (PATIENTS WITH TUBULAR PROTEINURIA)
RX MEDLINE=81184038; PubMed=6164372;
RA Takagi T.; Takagi K.; Kawai T.;

RT "Complete amino acid sequence of human alpha 1-microglobulin."
 RL Biochem. Res. Commun. 98:997-1001 (1981).
 RN [8]

RP SEQUENCE OF 206-350

RX MEDLINE-85225968; PubMed-2408638;
RA Reisinger P., Hochstrasser K., Albrecht G.J., Lempart K., Sailer J.P.;
RT "Human inter-alpha-trypsin inhibitor: localization of the kunitz-type
RT domains in the N-terminal part of the molecule and their release by a
RT trypsin-like proteinase.";
RL Biol. Chem. Hoppe-Seyler 366:479-483(1985).
RN [9]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-852074265; PubMed-6171497;
RA Hochstrasser K., Schoenberger O.L., Rossmann I., Wachter E.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RT the inter-alpha-trypsin inhibitor, V. Attachments of carbohydrates in
RT the human urinary trypsin inhibitor isolated by affinity
RT chromatography.";
RL Hoppe-Seyler's Z. Physiol. Chem. 362:1357-1362(1981).
RN [10]
RP INHIBITORY SITE.
RX MEDLINE-85225940; PubMed-3890890;
RA Moril M., Travis J.;
RT "The reactive site of human inter-alpha-trypsin inhibitor is in the
RT amino-terminal half of the protein.";
RL Biol. Chem. Hoppe-Seyler 366:19-21(1985).
RN [11]
RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE-90306345; PubMed-1694784;
RA Escaribano J., Lopez-Otin C., Hjerpe A., Grubb A., Mendez E.;
RT "Location and characterization of the three carbohydrate prosthetic
RT groups of human protein HC.";
RL FEBS Lett. 266:167-170(1990).
RN [12]
RP CHROMOPHORE.
RC TISSUE-Urine;
RX MEDLINE-91340714; PubMed-1714898;
RA Escaribano J., Grubb A., Calero M., Mendez E.;
RT "The protein HC chromophore is linked to the cysteine residue at
RT position 34 of the polypeptide chain by a reduction-resistant bond
RT and causes the charge heterogeneity of protein HC.";
RL J. Biol. Chem. 266:15758-15763(1991).
RN [13]
RP SEQUENCE OF 206-219, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
RC TISSUE-Plasma;
RX MEDLINE-94229087; PubMed-7513643;
RA Morelle M., Capon C., Balduyck M., Sautiere P., Kouach M.,
RA Michalski C., Fournet B., Mizon J.;
RT "Chondroitin sulphate covalently cross-links the three polypeptide
RT chains of inter-alpha-trypsin inhibitor.";
RL Eur. J. Biochem. 221:881-888(1994).
RN [14]
RP SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC2.
RX MEDLINE-9332026; PubMed-7682553;
RA Englund J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
RA Pizzo S.V., Hefta S.A.;
RT "Presence of the protein-glycosaminoglycan-protein covalent cross-link
RT in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
RT 2/bikunin.";
RL J. Biol. Chem. 268:8711-8716(1993).
RN [15]
RP SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC3.
RX MEDLINE-91093267; PubMed-1898736;
RA Englund J.J., Salvesen G., Hefta S.A., Thøgersen I.B.,
RA Rutherford S., Pizzo S.V.;
RT "Chondroitin 4-sulfate covalently cross-links the chains of the human
RT blood protein pre-alpha-inhibitor.";
RL J. Biol. Chem. 266:747-751(1991).
RN [16]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 230-339.
RX MEDLINE-98227321; PubMed-9566199;
RA Xu Y., Carr P.D., Guss J.M., Ollis D.L.;
RT "The crystal structure of bikunin from the inter-alpha-inhibitor
RT complex: a serine proteinase inhibitor with two kunitz domains.";
RL J. Mol. Biol. 276:955-966(1998).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT

CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN.
CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LISOSOMAL GRANULOCYTIC
CC ELASTASE. ADDITIONAL PROTEOLYTIC PROCESSING IN THE KIDNEY AND/OR
CC URINE CAN PRODUCE FURTHER AMINO- AND CARBOXYL-END MODIFICATIONS
CC IN ITS SEQUENCE.
CC -1- SUBUNIT: INTER-ALPHA-TRYPsin INHIBITOR CONSIST OF A LIGHT CHAIN
CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
CC -1- PTM: ADDITION OF GLYCOSAMINOGLYCAN CHONDROITIN SULFATE, ALLOWS
CC CROSS-LINKING BETWEEN THE DIFFERENT COMPONENTS.
CC -1- MISCELLANEOUS: IN VITRO, THE FIRST TWELVE RESIDUES OF THE AMINO
CC END OF THE INHIBITOR APPEAR TO HAVE A REACTIVE SITE CAPABLE OF
CC INHIBITING THE ACTIVITY OF A NUMBER OF ENZYMES. ITS IN VIVO
CC FUNCTION IS NOT KNOWN.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X54816; CAA38585.1; -
CC EMBL: X54817; CAA38585.1; JOINED.
CC EMBL: X54818; CAA38585.1; JOINED.
CC EMBL: X04225; CAA27803.1; -
CC EMBL: M88249; AAA59196.1; -
CC EMBL: M88165; AAA59196.1; JOINED.
CC EMBL: M88243; AAA59196.1; JOINED.
CC EMBL: M88244; AAA59196.1; JOINED.
CC EMBL: M88246; AAA59196.1; JOINED.
CC EMBL: M88247; AAA59196.1; JOINED.
CC EMBL: X04494; CAA28182.1; -
CC EMBL: X54817; CAA38586.1; -
CC PIR: A03217; HCHU
CC PIR: A25303; A25303.
CC PIR: S13433; S13433.
CC PIR: S10717; S10717.
CC PDB: 1B1K; 16-MAR-99.
CC DR SWISS-2DPAGE: P02760; HUMAN.
CC MIM: 176870; -
CC InterPro: IPR002223; Kunitz_BPTI.
CC InterPro: IPR002345; Lipocalin.
CC InterPro: IPR000566; Lipocalin_cytfabp.
CC Pfam: PF00014; Kunitz_BPTI; 2.
CC Pfam: PF00061; lipocalin; 1.
CC PRINTS: PR00179; LIPOCALIN.
CC PRINTS: PR00759; BASICPTASE.
CC SMART: SM00131; KU; 2.
CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
CC PROSITE: PS00279; BPTI_KUNITZ_2; 2.
CC PROSITE: PS00213; LIPOCALIN; 1.
CC KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
Query Match 7.5%; Score 213; DB 1; Length 352;
Best Local Similarity 40.2%; Pred. No. 7.4e-07;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;
OY 274 CIGPISPHLVNHYDPRQRCGMFPARGCDGARGETETACQACARPGDVCYPAVQ 333
DB 240 CMGMTSRYP-----YNGTSMACEFFOYGGCMGNNVTREKELQRTVA--ACNLPVIR 293
OY 334 GPCRGWEPWYASPLLOQCHPFYVGGCEGNGNNFHSRESCEDACPVP 380
DB 294 GPCRAFIQIMAFDAVKGCVLPFYGGCGGNGKFTSEKRECYGVP 340

RESULT	6			
AMBP_BOVIN	STANDARD;	PRT;	352 AA.	
ID	AMBP_BOVIN			
AC	P00978; P35420; Q28020;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-TRYPsin INHIBITOR LIGHT CHAIN (ITIL-IC) (BIXKININ) (HI-14) (COMULUS EXTRACELLULAR MATRIX STABILIZING FACTOR) (ESF)].			
GN	AMBP OR ITIL.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=96201710; PubMed=8611630;			
RA	Lindqvist A., Akerstrom B.;			
RT	"Bovine alpha 1-microglobulin/bikunin. Isolation and characterization			
RL	of liver cDNA and urinary alpha 1-microglobulin.";			
RL	Biochim. Biophys. Acta 1306:98-106(1996).			
RN	[2]			
RP	SEQUENCE OF 227-349.			
RX	MEDLINE=85225967; PubMed=2408637;			
RA	Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;			
RT	"Kunitz-type proteinase inhibitors derived by limited proteolysis of			
RL	trypsin-released inhibitors from horse and pig inter-alpha-trypsin			
RL	inhibitors.";			
RL	Biol. Chem. Hoppe-Seyler 366:473-478(1985).			
RN	[3]			
RP	SEQUENCE OF 227-348.			
RX	MEDLINE=8413807; PubMed=6199275;			
RA	Hochstrasser K., Wachter E.;			
RT	"Kunitz-type proteinase inhibitors derived by limited proteolysis of			
RL	the inter-alpha-trypsin inhibitor, VII. Determination of the			
RL	amino-acid sequence of the trypsin-released inhibitor from bovine			
RL	inter-alpha-trypsin inhibitor.";			
RL	Hoppe-Seyler's Z. Physiol. Chem. 364:1679-1687(1983).			
RP	[4]			
RP	SEQUENCE OF 206-219.			
RC	TISSUE=Fetal serum;			
RA	MEDLINE=92291130; PubMed=1376324;			
RT	Chen L., Mao S.J.T., Larsen W.J.;			
RT	"Identification of a factor in fetal bovine serum that stabilizes the			
RT	connulus extracellular matrix. A role for a member of the inter-alpha-			
RT	trypsin inhibitor family.";			
RL	J. Biol. Chem. 267:12380-12386(1992).			
RN	[5]			
RP	REACTIVE SITES.			
RX	MEDLINE=84133808; PubMed=6199276;			
RA	Hochstrasser K., Albrecht G.J., Schoenberger O.L., Wachter E.;			
RT	"Kunitz-type proteinase inhibitors derived by limited proteolysis of			
RL	the inter-alpha-trypsin inhibitor, VII. Characterization of the			
RL	bovine inhibitor as double-headed trypsin-elastase inhibitor.";			
RL	Hoppe-Seyler's Z. Physiol. Chem. 364:1689-1696(1983).			
CC	-1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL			
CC	FUNCTIONS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT			
CC	APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA			
CC	AND ALBUMIN.			
CC	-1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND			
CC	URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GLUCOCYTIC			
CC	ELASTASE.			
CC	-1- FUNCTION: MAY DIFFUSE INTO POLICULAR FLUID AFTER AN OVULATORY			
CC	STIMULUS TO ACT AS STRUCTURAL LINKER THAT ENSURE NORMAL COMULUS			
CC	EXPANSION, THROUGH STABILIZATION OF THE COMULUS EXTRACELLULAR			
CC	MATRIX THUS SUPPORTING THE PROCESS OF OVULATION.			
CC	-1- PRT: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO			

SEPARATELY FUNCTIONING PROTEINS.

-1- PTH: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.

-1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE LIPOCALIN FAMILY.

-1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/or_send_an_email_to_license@isb-sib.ch).

EMBL: U35642; AAB0759.1; -

PIR: A01209; TIRBO1.

DR HSSP; P10646; 1ADZ.

DR InterPro; IPR002223; Kunitz_BPTI.

DR InterPro; IPR002345; Lipocalin.

DR InterPro; IPR000566; Lipocalin_CYTBABP.

DR Pfam; PF00014; Kunitz_BPTI; 2.

DR Pfam; PF00061; lipocalin; 1.

DR PRINTS; PR00179; LIPOCALIN.

DR PRINTS; PR00759; BASICPTASE.

DR SMART; SM00131; KU; 2.

DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.

DR PROSITE; PS50279; BPTI_KUNITZ_2; 2.

DR PROSITE; PS00213; LIPOCALIN; 1.

DR KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat; lipocalin.

FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 203 ALPHA-1 MICROGLOBULIN.

FT CHAIN 206 352 INTER-ALPHA-TYPEPSIN INHIBITOR LIGHT CHAIN.

FT DOMAIN 231 281 BPTI/KUNITZ INHIBITOR 1.

FT BINDING 287 337 BPTI/KUNITZ INHIBITOR 2.

FT 53 53 CHROMOPHORE (BY SIMILARITY).

FT DISULFD 91 188 BY SIMILARITY.

FT DISULFD 231 281 BY SIMILARITY.

FT DISULFD 240 264 BY SIMILARITY.

FT DISULFD 256 277 BY SIMILARITY.

FT DISULFD 287 337 BY SIMILARITY.

FT DISULFD 326 320 BY SIMILARITY.

FT DISULFD 332 333 BY SIMILARITY.

FT ACT_SITE 241 242 INHIBITORY SITE (P1) (CHYMOTRYPSIN, ELASTASE).

FT ACT_SITE 297 298 INHIBITORY SITE (P1) (TRYPSIN).

FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .).

FT CONFLICT 209 209 T -> G (IN REF. 4).

FT CONFLICT 217 217 A -> D (IN REF. 4).

FT CONFLICT 268 268 G -> L (IN REF. 2 AND 3).

FT CONFLICT 274 274 E -> Q (IN REF. 2 AND 3).

FT CONFLICT 298 299 SY -> AF (IN REF. 2 AND 3).

FT CONFLICT 330 330 E -> Q (IN REF. 2 AND 3).

FT CONFLICT 346 346 E -> R (IN REF. 2 AND 3).

SQ SEQUENCE 352 AA; 39235 MW; ED31C5CA02E70B19 CRC64;

Query Match 7.4%; Score 210; DB 1; Length 352;

Best Local Similarity 24.2%; Pred. No. 1.le-06;

Matches 72; Conservative 34; Mismatches 93; Indels 98; Gaps 11;

156 LYSSSPQAVVGGTASLH-----CD-VSGRPVATVTEKSHQRENLMPD--QMYG 206

70 LTAGPTSKSEISVTN--HRRKGVESISG-----TYKTSADCKFLYHAKKNITMES 120

207 NVVVSIGVLVYKNRPEDAGLYCTARNAGILRADPLSVQ--REP----- 253

121 YVYHINVEYALFLTK-----LSRRHGPITVKKLYGNEPQLRESLREF 165

254 -----ARDAAISPAPAECLPDVQ----- 272

```

Db 166 REVALGVGIPEDALFTMPRGCEVPEGDUPVPTPLSRARAVLTQEBEGSGAGPVTNS 225
Oy 273 ----AC-----TGPTSPHLVLMHYDPORGGCMTFPARCGDGAARGFETYEACQOACARP 323
Db 226 KRADSCQLDYDSOGPCGLFKRYFVNGTSMACETFLYGGCGMGNNGNNFHSRECECLQTCRVA--T 283
Oy 324 GDAVCPLPANOGRGCRGMRWASPLLOQCHPFVYGGCEGNGNNFHSRECEADACPPV 380
Db 284 VECANLPIVGGPCRSYIQLMAFDVAKGKCVFRSYGGCKGNKMFSEKCEKEYCGIP 340

RESULT 7
IATR_HORSE STANDARD: PRT: 123 AA.
AC P04365:
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTER-ALPHA-TRYPSIN INHIBITOR (ITI) (HI-14) (INHIBITORY FRAGMENT OF ITI) (FRAGMENT).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RX MEDLINE=85225967; PubMed=2408637;
RA Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-alpha-trypsin inhibitor, X. The amino acid sequences of the trypsin-released inhibitors from horse and pig inter-alpha-trypsin inhibitors."
RT Biol. Chem. Hoppe-Seyler 366:473-478(1985).
CC -1- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS. WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND NOT AT ALL WITH PANCREATIC ELASTASE.
CC -1- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I.
CC INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
DR PIR: A01210; TIHOBI.
DR HSSP: P10646; IAD2.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT NON_TER 1 1
FT DOMAIN 5 55 BPTI/KUNITZ INHIBITOR 1.
FT DISULFID 61 111 BPTI/KUNITZ INHIBITOR 2.
FT DISULFID 14 55
FT DISULFID 30 51
FT DISULFID 61 111
FT DISULFID 70 94
FT DISULFID 86 107
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN, ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPSIN).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
FT NON_TER 123 123
SQ SEQUENCE 123 AA: 13510 MW: CELA9120774411D5 CRC64;

```

Query Match 7.3%; Score 208; DB 1; Length 123;
 Best Local Similarity 38.1%; Pred. No. 5.7e-07;
 Matches 40; Conservative 14; Mismatches 49; Indels 2; Gaps 1;

```

Oy 276 GPTSPHLVLMHYDPORGGCMTFPARCGDGAARGFETYEACQOACARGPDGACVLPVAGSP 335
Db 12 GPCLGMSIRYFNNGTSMACETFOYGGCLGNGNNFASQKECLQTCRVA--ACNLPVGGP 69
Oy 336 CRGMEPRWASPLLOQCHPFVYGGCEGNGNNFHSRECEADACPPV 380
Db 70 CRAFTLMAFDAAQCKVLFYGGCGMGNNGNNKFFSQKECEKEYCGIP 114

RESULT 8
IATR_SHEEP STANDARD: PRT: 123 AA.
AC P13371:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTER-ALPHA-TRYPSIN INHIBITOR (ITI) (GIR-14) (INHIBITORY FRAGMENT OF ITI) (FRAGMENT).
OS Ovis aries (Sheep), and Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940, 9925;
RN [1]
RP SEQUENCE.
RC SPECIES=Sheep.
RX MEDLINE=87299012; PubMed=2441725;
RA Rasp G., Hochstrasser K., Gerl C., Wachter E.;
RT "The amino acid sequence of the trypsin-released inhibitor from sheep inter-alpha-trypsin inhibitor."
RT Biol. Chem. Hoppe-Seyler 368:727-731(1987).
RN [2]
RP SEQUENCE.
RC SPECIES=C. hircus;
RX MEDLINE=90105540; PubMed=2481505;
RA Rasp G., Hochstrasser K., Gerl C., Wachter E.;
RT "Primary structure of a proteinase inhibitor released from goat serum inter-alpha-trypsin inhibitor."
RT Biochim. Biophys. Acta 999:335-337(1989).
CC -1- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS. WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND NOT AT ALL WITH PANCREATIC ELASTASE.
CC -1- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I.
CC INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
DR PIR: A29652; A29652.
DR HSSP: P10646; IAD2.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF00014; Kunitz_BPTI; 2.
DR SMART: SM00131; KU; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT NON_TER 1 1
FT DOMAIN 5 55 BPTI/KUNITZ INHIBITOR 1.
FT DISULFID 61 111 BPTI/KUNITZ INHIBITOR 2.
FT DISULFID 14 55
FT DISULFID 30 51
FT DISULFID 61 111
FT DISULFID 70 94
FT DISULFID 86 107
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN, ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPSIN).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
FT NON_TER 123 123
SQ SEQUENCE 123 AA: 13686 MW: 295038173722D2D1 CRC64;

```

```

Query Match          7.2%; Score 205; DB 1; Length 133;
Best Local Similarity 37.1%; Pred. No. 8.8e-07;
Matches 39; Conservative 15; Mismatches 49; Indels 2; Gaps 1;

OY 276 GPTSPHLYMHYDPQRCGGCTFPARGCCDCAARGFEYACQACARGPDACVAVGCP 335
DB 12 GTCGLGMRKRYFYNGTSMACETFYGGCGMGNNGNPFSEKCLQTCR--TVQACNLPVIRGP 69

OY 336 CCGWEPRWAYSPLLQCHPFFYGGCGEGNGNPFHRSCEADACVP 380
DB 70 CRAGIELMAFDVAKGKCVAFYGGCGNGNGNPFYSQKECEYGCIP 114

RESULT 9
TFPI_RABIT
ID TFPI_RABIT STANDARD; PRT; 300 AA.
AC P19761; Q28828;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)
DE (EPI).
GN TFPI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91057146; PubMed=2136251;
RA Wesselschmidt R.L., Girard T.J., Broze G.J. Jr.;
RT "cDNA sequence of rabbit lipoprotein-associated coagulation
RT inhibitor."
RL Nucleic Acids Res. 18:6440-6440(1990).
RN [2]
RP REVISIONS TO 72; 211 AND 218.
RC TISSUE=Liver;
RX MEDLINE=92335027; PubMed=1630940;
RA Wain-Cramer B.J., Broze G.J. Jr., Komives E.A.;
RT "cDNA sequence of rabbit tissue factor pathway inhibitor."
RL Nucleic Acids Res. 20:3548-3548(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=93276427; PubMed=8503123;
RA Beliaouaj A., Kuppaswamy M.N., Birkoft J.J., Bajaj S.P.;
RT "Revised cDNA sequence of rabbit tissue factor pathway inhibitor."
RL Thromb. Res. 69:547-553(1993).
CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
CC A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN
CC ANTITHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
CC LIPOPROTEINS IN PLASMA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54708; CAA38515.1; ALT_SEQ.
DR EMBL; S61902; AAB26836.1; -.
DR PIR; S12143; S12143.
DR HSSP; P10646; ITFX.
DR InterPro; IPR002223; Kunitz_BPTI.

```

```

DR Pfam; PF00014; Kunitz_BPTI; 3.
DR PRINTS; PRO0759; BASICPTASE.
DR SMART; SMO0131; KU; 3.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 3.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 3.
KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
KW signal.
FT SIGNAL 1 24
FT CHAIN 25 300
FT DOMAIN 50 100
FT 121 171
FT 213 263
FT DOMAIN
FT DISULFID 50 100
FT DISULFID 59 83
FT DISULFID 75 96
FT ACT_SITE 60 61
FT DISULFID 121 171
FT DISULFID 130 154
FT DISULFID 146 167
FT ACT_SITE 131 132
FT DISULFID 213 263
FT DISULFID 222 246
FT DISULFID 238 259
FT ACT_SITE 223 224
FT CARBOHYD 141 141
FT CARBOHYD 191 191
FT CARBOHYD 252 252
FT CONFLICT 31 31
FT CONFLICT 269 272
SQ SEQUENCE 300 AA; 34435 MW; A08DE36537708CA6 CRC64;

Query Match          7.1%; Score 204; DB 1; Length 300;
Best Local Similarity 31.1%; Pred. No. 2.4e-06;
Matches 50; Conservative 20; Mismatches 71; Indels 20; Gaps 3;

OY 238 GLTRADPEPLSVQREPAR-----DAAPSIPAPAECLPDVQACTGPTSPHVLVHYPQSG 292
DB 16 GLVPA--PVSSAAEDEFNTITDIRPLOKPHSHSCAMKVDGPCRAVTKREFEITLH 73
OY 293 GCMTPFARGCDGARGFEYACQACAR-----GPGDACYLPAVAGPCRGW 339
DB 74 QCEEFYGGCGEGNENRFESECKEKCARDYPKMTKTLFGQKGPFCFLIEDPGICRGY 133
OY 340 EPRWAYSPLLQCHPFFYGGCGEGNGNPFHRSCEADACVP 380
DB 134 ITRIFYNOSKQCEKRRFYGGCLGLNPFSELECKNTCENP 174

RESULT 10
AMP_MOUSE
ID AMP_MOUSE STANDARD; PRT; 349 AA.
AC 007456; Q61294;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AMP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKONIN) (HI-30)].
GN AMP OR ITIL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAUB/C; TISSUE=Liver;
RX MEDLINE=93363639; PubMed=7689339;
RA Chan P., Salier J.P.;
RT "Mouse alpha-1-microglobulin/bikunin precursor: cDNA analysis, gene
RT evolution and physical assignment of the gene next to the orosomucoid

```

RT locus.";
 RL Blochlm. Biophys. Acta 1174:195-200(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-Liver;
 RX MEDLINE-95189774; PubMed-7533761;
 RA Itoh H., Ide H., Kataoka H., Tomita M., Yoshihara H., Nawa Y.;
 RT "cDNA sequencing of mouse alpha 1-microglobulin/inter-alpha-trypsin
 inhibitor light chain and its expression in acute inflammation.";
 RJ. Biochem. 116:767-772(1994).
 RN [3]
 RP SEQUENCE OF 128-349 FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-Liver;
 RA Itoh H., Ide H., Yoshihara H., Nawa Y.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
 CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
 CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
 CC AND ALBUMIN (BY SIMILARITY).
 CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
 CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
 CC ELASTASE (BY SIMILARITY).
 CC -1- SUBUNIT: INTER-ALPHA-TRYPsin INHIBITOR CONSIST OF A LIGHT CHAIN
 CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
 CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
 CC SEPARATELY FUNCTIONING PROTEINS.
 CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X68680; CAA48640.1; -
 CC EMBL: D28812; BAA05973.1; -
 CC HSSP: P12111; 1KUN.
 CC MGD: MGI:88002; Ambp.
 CC DR InterPro: IPR002223; Kunitz_BPTI.
 CC DR InterPro: IPR002345; Lipocalin.
 CC DR InterPro: IPR003566; Lipocalin_CYFABP.
 CC DR Pfam: PF00061; Kunitz_BPTI; 2.
 CC DR Pfam: PF00061; Lipocalin; 1.
 CC DR PRINTS: PR00179; LIPOCALIN.
 CC DR PRINTS: PR00759; BASICPTASE.
 CC DR SMART: SM00331; KU; 2.
 CC DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 CC DR PROSITE: PS0279; BPTI_KUNITZ_2; 2.
 CC DR PROSITE: PS00213; LIPOCALIN; 1.
 CC KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
 CC Lipocalin.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 202
 CC FT CHAIN 205 349
 CC FT
 CC FT DOMAIN 230 280
 CC FT DOMAIN 285 336
 CC FT BINDING 52 52
 CC FT BINDING 52 52
 CC FT DISULFID 90 187
 CC FT DISULFID 230 280
 CC FT DISULFID 239 263
 CC FT DISULFID 255 276
 CC FT DISULFID 286 336
 CC FT DISULFID 295 319
 CC FT DISULFID 311 332
 CC FT CARBOHYD 33 33
 CC FT CARBOHYD 114 114
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 114

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT ACT_SITE 240 241 INHIBITOR SITE (PI) (CHYMOTRYPSIN,
 FT ELASTASE) (BY SIMILARITY).
 FT ACT_SITE 296 297 INHIBITOR SITE (PI) (TRYPSIN) (BY
 FT CONFLICT 65 65 SIMILARITY).
 FT CONFLICT 65 65 Q -> S (IN REF. 2).
 SQ SEQUENCE 349 AA; 39070 MW; CE4D9FC737DA80B CRC64;
 Query Match 7.18; Score 203; DB 1; Length 349;
 Best Local Similarity 40.68; Pred. No. 3.1e-06;
 Matches 39; Conservative 13; Mismatches 42; Indels 2; Gaps 1;
 QY 285 MHNDPQGGCMTPFARCCDGAARFETFEACQACARGPDACVLPVAGRCRWGRPWA 344
 DB 246 YYNGASMAETETGYGCLGNGNNEFSEKDLQCRITA-ACNLPVQGCRAFIKMA 303
 QY 345 YSPLLQCHPEFVYGCCEGNGNNEFHSRSCEDACVP 380
 DB 304 FDAAGKCIQPHYGCCKNGKNGKFESEKCEKCYGVP 339
 RESULT 11
 TFPI_HUMAN
 ID TFPI_HUMAN STANDARD; PRT; 304 AA.
 AC P10646;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)
 DE (TFPI).
 GN TFPI OR TFPI1 OR LACI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-88198127; PubMed-2452157;
 RA Wan T.-C., Kretzmer K.K., Girard T.J., Miletich J.P., Broze G.J. Jr.;
 RT "Cloning and characterization of a cDNA coding for the lipoprotein
 RT associated coagulation inhibitor shows that it consists of three
 RT tandem Kunitz-type inhibitor domains.";
 RL J Biol. Chem. 263:6001-6004(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91129227; PubMed-1993173;
 RA van der Logt C.P.E., Reitsma P.H., Bertina R.M.;
 RT "Intron-exon organization of the human gene coding for the
 RT lipoprotein-associated coagulation inhibitor: the factor Xa dependent
 RT inhibitor of the extrinsic pathway of coagulation.";
 RL Biochemistry 30:1571-1577(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91161593; PubMed-2002045;
 RA Girard T.J., Eddy R., Wesselschmidt R.L., Macphail L.A.;
 RA Likert K.W., Byers W.G., Shows T.B., Broze G.J. Jr.;
 RT "Structure of the human lipoprotein-associated coagulation inhibitor
 RT gene. Intron/exon gene organization and localization of the gene to
 RT chromosome 2.";
 RL J Biol. Chem. 266:5036-5041(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89388722; PubMed-2781520;
 RA Girard T.J., Warren L.A., Novotny W.F., Bejcek B.E., Miletich J.P.,
 RA Broze G.J. Jr.;
 RT "Identification of the 1.4 kb and 4.0 kb messages for the lipoprotein
 RT associated coagulation inhibitor and expression of the encoded
 RT protein.";
 RL Thromb. Res. 55:37-50(1989).
 RN [5]
 RP SEQUENCE OF 29-50.

RX MEDLINE-90036996; PubMed-2533722;
 RA Novotny W.F., Girard T.J., Milelich J.P., Broze G.J. Jr.;
 RT "Purification and characterization of the lipoprotein-associated
 RT coagulation inhibitor from human plasma";
 RL J. Biol. Chem. 264:18832-18837(1989).
 RN [6]
 RP INHIBITORY SITES.
 RX MEDLINE-89181950; PubMed-2927510;
 RA Girard T.J., Warren L.A., Novotny W.F., Likert K.M., Brown S.G.,
 RA Milelich J.P., Broze G.J. Jr.;
 RT "Functional significance of the Kunitz-type inhibitory domains of
 RT lipoprotein-associated coagulation inhibitor";
 RL Nature 338:518-520(1989).
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE-96224851; PubMed-8639592;
 RA Nakahara Y., Miyata T., Hamuro T., Funatsu A., Miyagi M.,
 RA Tsunasawa S., Kato H.;
 RT "Amino acid sequence and carbohydrate structure of a recombinant
 RT human tissue factor pathway inhibitor expressed in Chinese hamster
 RT ovary cells: one N- and two O-linked carbohydrate chains are located
 RT between Kunitz domains 2 and 3 and one N-linked carbohydrate chain is
 RT in Kunitz domain 2";
 RL Biochemistry 35:6450-6459(1996).
 RN [8]
 RP REVIEW.
 RX MEDLINE-91104709; PubMed-2271516;
 RA Broze G.J. Jr., Girard T.J., Novotny W.F.;
 RT "Regulation of coagulation by a multivalent Kunitz-type inhibitor";
 RL Biochemistry 29:7539-7546(1990).
 RN [9]
 RP STRUCTURE BY NMR OF 121-182.
 RX MEDLINE-97342711; PubMed-9199408;
 RA Burgering M.J., Orbons L.P., van der Doelen A., Mulders J.,
 RA Theunissen H.J., Groothuis P.D., Bode W., Huber R., Stubbs M.T.;
 RT "The second Kunitz domain of human tissue factor pathway inhibitor:
 RT cloning, structure determination and interaction with factor Xa";
 RL J. Mol. Biol. 269:395-407(1997).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
 CC A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN
 CC ANTIHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 CC LIPOPROTEINS IN PLASMA.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: MOSTLY IN ENDOTHELIAL CELLS.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- PTM: O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 CC EMBL; J03325; AAA52022.1; -;
 DR EMBL; M58650; AAA59480.1; -;
 DR EMBL; M58644; AAA59480.1; JOINED.
 DR EMBL; M58645; AAA59480.1; JOINED.
 DR EMBL; M58646; AAA59480.1; JOINED.
 DR EMBL; M58647; AAA59480.1; JOINED.
 DR EMBL; M58648; AAA59480.1; JOINED.
 DR EMBL; M58649; AAA59480.1; JOINED.
 DR EMBL; M59499; AAA59526.1; -;
 DR EMBL; M59493; AAA59526.1; JOINED.
 DR EMBL; M59494; AAA59526.1; JOINED.
 DR EMBL; M59495; AAA59526.1; JOINED.
 DR EMBL; M59496; AAA59526.1; JOINED.
 DR EMBL; M59497; AAA59526.1; JOINED.
 DR EMBL; M59498; AAA59526.1; JOINED.
 DR EMBL; M59499; AAA59526.1; JOINED.
 DR PIR; A28650; THHGK.

DR PIR; A34315; A34315.
 DR PIR; A60433; A60433.
 DR PIR; S03903; S03903.
 DR PDB; 1AD2; 25-FEB-98.
 DR PDB; 1TFX; 21-JAN-98.
 DR GlycosultEDB; P10646; -.
 DR MIM; 152310; -.
 DR InterPro; IP002223; Kunitz-BPTI.
 DR Pfam; PF00014; Kunitz-BPTI; 3.
 DR PRINTS; PR00759; BASICTPASE.
 DR SMART; SM00131; KU; 3.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 3.
 KW Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 KW Signal; 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 304
 FT DOMAIN 54 104
 FT DOMAIN 125 175
 FT DOMAIN 217 267
 FT DISULFID 54 104
 FT DISULFID 63 87
 FT DISULFID 79 100
 FT ACT_SITE 64 65
 FT DISULFID 125 175
 FT DISULFID 134 158
 FT DISULFID 150 171
 FT ACT_SITE 135 136
 FT DISULFID 217 267
 FT DISULFID 226 250
 FT ACT_SITE 227 228
 FT CARBOHD 145 145
 FT CARBOHD 202 202
 FT CARBOHD 203 203
 FT CARBOHD 195 195
 FT CARBOHD 64 64
 FT MUTAGEN 135 135
 FT MUTAGEN 227 227
 FT MUTAGEN 227 227
 SQ SEQUENCE 304 AA; 35015 MW; 5281E32B758B4FE CRC64;

 Query Match 6.8%; Score 194.5; DB 1; Length 304;
 Best Local Similarity 27.3%; Pred. No. 9.4e-06;
 Matches 50; Conservative 20; Mismatches 58; Indels 55; Gaps 5;

 QY 231 CTARNAAGILRADPLSVOREPARDAPSIAPAECL--PDVQACTGPTSPHLVLMHYD 288
 DB 104 CTBDNANRIK-----TLQQE-----KPDFCLEEDDPGICRG---YIRPYFN 144
 QY 289 PORGCGMFPFARGCGCARGPETYACQACARGG----- 324
 DB 145 NOTKOCERFKYGGCGGNKNNEETLECKNICEDEGNGFQVDNYGTQLNAVNSLTPQSTK 204
 QY 325 -----DACVLPAVGPGRGWEPRMAYSPILQQCHPPYVGCGEGNGNNFSRECE 374
 DB 205 VPSLFEFFHGPSKCLTPADRGICLRANENRFTYNSYIGKCRPKRYSCGGSENNFTSKQCL 264
 QY 375 DAC 377
 DB 265 RAC 267

 RESULT 12
 SPT2_HUMAN STANDARD: PRT; 252 AA.
 AC 043291; 000271; 014895;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR

DE ACTIVATOR INHIBITOR TYPE 2) (HA1-2) (PLACENTAL BIKUNIN).
GN SPINT2 OR HA12 OR KOP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98010584; PubMed=9346890;
RA Kawaguchi T., Qin L., Shimomura T., Kondo J., Matsumoto K., Denda K.,
RA Kitamura N.;
RT "Purification and cloning of hepatocyte growth factor activator
RT inhibitor type 2, a Kunitz-type serine protease inhibitor.";
RL J. Biol. Chem. 272:27558-27564(1997).
RP [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-74.
RC TISSUE=placenta;
RX MEDLINE=97273732; PubMed=9115294;
RA Marlor C.W., Delaria K.A., Davis G., Muller D.K., Greve J.M.,
RA Tamburini P.P.;
RT "Identification and cloning of human placental bikunin, a novel serine
RT protease inhibitor containing two Kunitz domains.";
RL J. Biol. Chem. 272:12202-12208(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=pancreatic cancer;
RX MEDLINE=98094245; PubMed=9434156;
RA Mueller-Pillaach F., Waltrapp C., Bartels K., Varga G., Fries H.,
RA Buechler M., Adler G., Gress T.M.;
RT "Cloning of a new Kunitz-type protease inhibitor with a putative
RT transmembrane domain overexpressed in pancreatic cancer.";
RL Biochim. Biophys. Acta 1395:88-95(1998).
CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR. ALSO INHIBITS PLASMIN,
CC PLASMA AND TISSUE KALLIKREIN, AND FACTOR XII.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, KIDNEY, PANCREAS,
CC PROSTATE, TESTIS, THYMUS, AND TRACHEA.
CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AB006534; BAA25024.1; -
CC DR EMBL: U78095; AAC02781.1; -
CC DR EMBL: AF027205; MAB84031.1; -
CC DR HSP: P05067; ITAW.
CC DR MIM: 605124; -
CC DR InterPro: IPR002223; Kunitz_BPTI.
CC DR Pfam: PF00014; Kunitz_BPTI_2.
CC DR PRINTS: PR00759; BASICPTASE.
CC DR SMART: SM00131; KU; 2.
CC DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
CC DR PROSITE: PS50279; BPTI_KUNITZ_2; 2.
CC KW serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;
CC Signal.
FT SIGNAL 1 27
FT CHAIN 28 252
FT DOMAIN 197 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.
FT TRANSMEM 198 218 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 219 252 POTENTIAL.
FT DOMAIN 219 252 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 88 BPTI/KUNITZ INHIBITOR 1.
FT DOMAIN 133 183 BPTI/KUNITZ INHIBITOR 2.
FT DISULFID 38 88 BY SIMILARITY.
FT DISULFID 47 71 BY SIMILARITY.
FT DISULFID 63 84 BY SIMILARITY.
FT ACT_SITE 48 49 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 133 183 BY SIMILARITY.

FT DISULFID 142 166 BY SIMILARITY.
FT DISULFID 158 179 BY SIMILARITY.
FT ACT_SITE 143 144 REACTIVE BOND (BY SIMILARITY).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 3 3 Q -> H (IN REF. 3).
FT CONFLICT 11 11 R -> P (IN REF. 1).
FT CONFLICT 53 53 R -> K (IN REF. 3).
FT CONFLICT 240 240 D -> H (IN REF. 3).
SQ SEQUENCE 252 AA; 28228 MW; A7D3360CDECBAB2B CQC64;
Query Match 6.7%; Score 192; DB 1; Length 252;
Best Local Similarity 28.5%; Pred No. 1; le-05;
Matches 45; Conservative 13; Mismatches 54; Indels 46; Gaps 4;
QY 267 CLPDVQACTGPTSPHVLVMDHPDQRCMTFFPARGCDGARGFETVACQACA----- 320
DB 38 CL--VSKVVGRCRASPMRWVNTDSCQLFYVGGDGSNNVLTKECLKCATVTENA 95
QY 321 -----RGFGDA-----CULPVGCPCRGHEPRWVAYSPLL 349
DB 96 TGDLATSRNADSSVSPAPRRQDSEDSMDMEYEECTANAVTGPCRASFPWYEDVER 155
QY 350 QQCFFVYGGCGEGNGNHFHRESCEDAC-----DVP 380
DB 156 NSCNFTYGGCGRGNKNSYSEACMLRCFRQENPLP 193
RESULT 13
SPRT2_MOUSE STANDARD; PRT; 252 AA.
AC Q9WU03; Q9WU04; Q9WU05;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE KUNITZ-TYPE PROTEASE INHIBITOR 2 PRECURSOR (HEPATOCYTE GROWTH FACTOR
DE ACTIVATOR INHIBITOR TYPE 2) (HA1-2).
GN SPINT2 OR HA12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BA1B/C;
RX MEDLINE=99160423; PubMed=10049781;
RA Itoh H., Kataoka H., Hamasuna R., Kitamura N., Koono M.;
RT "Hepatocyte growth factor activator inhibitor type 2 lacking the first
RT Kunitz-type serine proteinase inhibitor domain is a predominant
RT product in mouse but not in human.";
RL Biochem. Biophys. Res. Commun. 253:740-748(1999).
CC -1- FUNCTION: INHIBITOR OF HGF ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: ISOFORM 2 IS MORE PREDOMINANTLY EXPRESSED THAN
CC ISOFORM 1.
CC -1- DOMAIN: THIS INHIBITOR CONTAINS TWO INHIBITORY DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 BPTI/KUNITZ INHIBITOR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF099016; AAD22173.1; -
CC DR EMBL: AF099019; AAD22173.1; -
CC DR EMBL: AF099020; AAD22174.1; -
CC HSP: P05067; ITAW.

DR MGD: MGI:1338031; Spin12.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI. 2.
 DR PRINTS: PR00759; BASICTPASE.
 DR SMART: SM00131; KU. 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 2.
 DR Serine protease inhibitor; Repeat; Glycoprotein; Transmembrane;
 KW Signal; Alternative splicing.
 FT SIGNAL 1 27
 FT CHAIN 28 252
 FT DOMAIN 28 252 KUNITZ-TYPE PROTEASE INHIBITOR 2.
 FT TRANSMEM 198 218
 FT TRANSMEM 198 218
 FT DOMAIN 219 252
 FT DOMAIN 38 88
 FT DOMAIN 133 183
 FT DISULFID 38 88
 FT DISULFID 47 71
 FT DISULFID 47 71
 FT DISULFID 63 84
 FT ACT_SITE 48 49
 FT DISULFID 133 183
 FT DISULFID 142 166
 FT DISULFID 158 179
 FT ACT_SITE 143 144
 FT CARBOHYD 57 57
 FT CARBOHYD 94 94
 FT VARSPLIC 37 93
 FT VARSPLIC 114 128
 FT VARSPLIC 129 252
 FT SEQUENCE 252 AA; 27914 MW; B2F4B86924D4F8F CRC64;

Query Match 6.7%; Score 192; DB 1; Length 252;
 Best Local Similarity 26.1%; Pred. No. 1.1e-05;
 Matches 54; Conservative 20; Mismatches 87; Indels 46; Gaps 5;
 QY 231 CTARAAALLR--ADFLSYVQREPARDAAPSIPAPAECLPDVQACTGPTSPHLVLMHYD 288
 DB 5 CELRGRLALVAVSLISL-----GAOVAARELDVHESC--GVSKVVGKCRASIPRMWYN 57
 QY 289 PORGCMTPARGCDGARGFETYEACQACA-----RGGDA----- 326
 DB 58 ITDSCOPFVVGCGEGNGNNVQSKKECLDKACAGTENTTDMARRNGADSSVLSVPRKQ 117
 QY 327 -----CYLPAVQGPCRGWEPWAYSPLLOCHPVPYVGGCGGNGNNHRSRE 371
 DB 118 SAEDLSAEIFMYEEYCVPAKATGCPRAAFPRMYDYDEKNSCISFLYGGCRGNKNSYLSQE 177
 QY 372 SCEACVPPRPPCRACRLRSKLLSL 398
 DB 178 ACMQHCSGKQMHPLPLGLKAVLVVL 204
 RESULT 14
 TEPL_MOUSE
 ID TEPL_MOUSE STANDARD: PRT; 306 AA.
 AC 054819;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-
 DE ASSOCIATED COAGULATION INHIBITOR) (UACI) (EXTRINSIC PATHWAY INHIBITOR)
 DE (TEPI).
 GN TEPI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=98152575; PubMed=9493581;

RA Chang J.-Y., Monroe D.M., Oliver J.A., Liles D.K., Roberts H.R.;
 RT "Cloning, expression, and characterization of mouse tissue factor
 RT pathway inhibitor (TEPI)."
 RL Thromb. Haemost. 79:306-309(1998).
 CC -1- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT
 CC WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING
 CC A QUATERNARY X(A)/UACI/VII(A)/TF COMPLEX. IT POSSESSES AN
 CC ANTIHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH
 CC LIPOPROTEINS IN PLASMA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 BPTI/KUNITZ INHIBITOR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF004833; AAC40035.1; -.
 DR HSSP: P10646; ITEX.
 DR MGD: MGI:1095418; Tfpi.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR Pfam: PF00014; Kunitz_BPTI. 3.
 DR PRINTS: PR00759; BASICTPASE.
 DR SMART: SM00131; KU. 3.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 3.
 DR Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;
 KW Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 306
 FT DOMAIN 50 100
 FT DOMAIN 121 171
 FT DOMAIN 225 275
 FT DISULFID 50 100
 FT DISULFID 59 83
 FT DISULFID 75 96
 FT ACT_SITE 60 61
 FT DISULFID 121 171
 FT DISULFID 130 154
 FT DISULFID 146 167
 FT ACT_SITE 131 132
 FT DISULFID 225 275
 FT DISULFID 234 258
 FT DISULFID 250 271
 FT ACT_SITE 235 236
 FT CARBOHYD 141 141
 FT CARBOHYD 254 254
 FT CARBOHYD 264 264
 FT CARBOHYD 282 282
 FT SEQUENCE 306 AA; 34987 MW; D3EA3297E4B6A359 CRC64;

Query Match 6.5%; Score 186.5; DB 1; Length 306;
 Best Local Similarity 31.8%; Pred. No. 3e-05;
 Matches 41; Conservative 17; Mismatches 54; Indels 17; Gaps 2;
 QY 271 VQACTGPTSPHLVLMHYDPORGCMTPARGCDGARGFETYEACQACARGP----- 324
 DB 52 MKADDGPKAMIRSYFNMTHTQCEETVGGCGNGNRRPDTLECKKTC--PGYEKTV 109
 QY 325 -----DACYLPAVQGPCRGWEPWAYSPLLOCHPVPYVGGCGGNGNNHRSRSCED 375
 DB 110 KAASGAERPDFCFLEBDPGICRGYMKRYLYNNQKOCERFVVGCGLGNNHNETLECKR 169
 QY 376 ACPVPRTP 384
 DB 170 ICENPVHSP 178

Db 859 RFAFDAEAQDCRAFTYGGCGGNGNNFATMQECRRCVMAMKKSPVATCEADIEVGECAGV 91

C:Species: Sus scrofa domestica (domestic pig)
N,Airte name: Dikahin, III, PI-14 (Inhib)

C>Date: 30-Jun-1987 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: S11066; S13493; A01208
R:Gebhard, W.; Schetlmueller, T.; Vetr, H.; Wachter, E.; Hochstrasser, K.
FEBS Lett. 269, 32-36, 1990
A:Title: Complementary DNA and deduced amino acid sequences of porcine alpha1-microglobulin
A:Reference number: S11066; MUID:90353595
A:Accession: S11066
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-337 <GB>
A:Cross-references: EMBL:X53685; NID:g1877; PIDN:CAA37725.1; PID:g1878
R:Tavakkoli, A.
Biochim. Biophys. Acta 1088, 47-56, 1991
A:Title: Molecular cloning of porcine alpha(1)-microglobulin/HI-30 reveals developmental
A:Reference number: S13493; MUID:91113729
A:Accession: S13493
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'M', '3-48', 'M', '50-337' <TAV>
A:Cross-references: GB:X52087; NID:g1881; PIDN:CAA36306.1; PID:g1882
A:Note: The authors translated the codon GTG for residue 2 as a Met initiation codon
R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Reisinger, P.
Biol. Chem. Hoppe-Seyler 366, 473-478, 1985
A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-
A:Reference number: A90685; MUID:85225967
A:Accession: A01208
A:Molecule type: protein
A:Residues: 212-258, 'Q', '260-269', 'S', '271-277', 'O', '279-282', 'A', '284', 'IR', '287-292', 'A', '294-310'
C:Comment: This inhibitory fragment, released from native ITI after limited proteolysis
first domain interacts weakly with PMN-granulocytic elastase and not at all with pancreas
C:Comment: The amino acid at position P2' (228-Met) appears to determine the specificity
and elastase; those with leucine interact strongly.
C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocalin
C:Keywords: duplication; glycoprotein; plasma; serine proteinase inhibitor
F:20-173/Domain: lipocalin homology <LRP>
F:216-266/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
F:272-322/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
F:216-266, 225-249, 241-262, 272-332, 281-305, 297-318/Dsulfide bonds: #status predicted
F:226/Inhibitory site: Leu (chymotrypsin, elastase) #status predicted
F:235/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:282/Inhibitory site: Arg (trypsin) #status predicted

Query Match 7.5%; Score 214; DB: 1; Length 337;
Best Local Similarity 39.0%; Pred. No. 6e-06;
Matches 41; Conservative 15; Mismatches 47; Indels 2; Gaps 1;

QY 276 GPTSHLVLMHDPQRCGMPFARCGDCAARGFETFEACQOACARGPDACVLPVAGCP 335
DB 223 GPCGLMKIRFYNGSSMACETPHYGGCMGNGNMFVSEKCLQTCR--TYEACSLPIVSGP 280
QY 336 CGWEPFMAVSPFLQOCHPFYVGGCEGNGNMFHRESCEDACVPV 380
DB 281 CRGFQLMNFAVDAGKCVLFNYGGCGGNGNGFSEKECKEYGVGP 325

RESULT 5
S11069
alpha-1-microglobulin/inter-alpha-trypsin inhibitor light chain precursor - rat
N:Alternate names: acid-stable proteinase inhibitor; bikunin; trypstatin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C:Accession: S21089; A53056; A25935; A31890; A61633
R:Lindqvist, A.; Bratt, T.; Alsterl, M.; Kaster, W.; Akerstrom, B.
Biochim. Biophys. Acta 1130, 63-67, 1992
A:Title: Rat alpha(1)-microglobulin: co-expression in liver with the light chain of inter-
A:Reference number: S21089; MUID:92182014
A:Accession: S21089
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-349 <LIN>
A:Cross-references: GB:S87544; NID:9247162; PIDN:AAB21782.1; PID:9247163
R:Itch, H.; Ide, H.; Ishikawa, N.; Nawa, Y.

J. Biol. Chem. 269, 3818-3822, 1994
A:Title: Mast cell protease inhibitor, trypstatin, is a fragment of inter-alpha-tryps
A:Reference number: A53056; MUID:94148892
A:Accession: A53056
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 285-341 <ITN>
R:Kaster, W.; Bjorck, L.; Akerstrom, B.
J. Biol. Chem. 261, 15070-15074, 1986
A:Title: Developmental and tissue-specific expression of alpha-1-microglobulin mRNA
A:Reference number: A25935; MUID:87033744
A:Accession: A25935
A:Molecule type: protein
A:Residues: 141, 'A', '143-195' <KAS>
R:Kido, H.; Yokogoshi, Y.; Katunuma, N.
J. Biol. Chem. 263, 18104-18107, 1988
A:Title: Kunitz-type protease inhibitor found in rat mast cells. Purification, proper
A:Reference number: A31890; MUID:89053978
A:Accession: A31890
A:Molecule type: protein
A:Residues: 283-301, 'L', '303-322', 'N', '324-329', 'PK', '332-333', 'W', '335-343' <KID>
R:Sugik, M.; Maruyama, M.; Yoshida, E.; Sumi, H.; Mihara, H.
Inflammation 15, 281-289, 1991
A:Title: Acid-stable protease inhibitor in chronic phase of carrageenin-induced infla
A:Reference number: A61633; MUID:92120777
A:Accession: A61633
A:Molecule type: protein
A:Residues: 205-213, 'X', '215-229', 'N', '231-232', 'K', '234-238' <SUG>
C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocalin
C:Keywords: chondroitin sulfate proteoglycan; chromoprotein; glycoprotein; plasma; se
F:1-19/Domain: signal sequence #status predicted <STG>
F:20-202/Product: alpha-1-microglobulin #status predicted <ALN>
F:34-187/Domain: lipocalin homology <LRP>
F:205-344/Product: inter-alpha-trypsin inhibitor #status experimental <IAI>
F:230-280/Domain: animal Kunitz-type proteinase inhibitor homology <BP1>
F:286-336/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>
F:24/Binding site: carbohydrate (Thr) (covalent) #status predicted
F:52/Cross-link: alpha-1-microglobulin-ig alpha complex chromophore (Cys) (interchain
F:114, 233/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:214/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental
F:296/Inhibitory site: Arg (trypsin) #status predicted

Query Match 7.5%; Score 214; DB: 2; Length 349;
Best Local Similarity 40.6%; Pred. No. 6.2e-06;
Matches 39; Conservative 14; Mismatches 41; Indels 2; Gaps 1;

QY 285 MHNDPQRCGMPFARCGDCAARGFETFEACQOACARGPDACVLPVAGCP 344
DB 246 YYINGSMACETPHYGGCMGNGNMFASKECLQTCRTIA--ACNLPYVQPCRAFLMA 303
QY 345 YSPFLQOCHPFYVGGCEGNGNMFHRESCEDACVPV 380
DB 304 FDAAGKCIQFIYGGCKGNGNMFSEKECKEYGVGP 339

RESULT 6
H0HU
alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor [validated] - human
N:Alternate names: bikunin; complex-forming glycoprotein heterogeneous in charge (HC)
rich protein
N:Contains: alpha-1-microglobulin (protein HC); inter-alpha-trypsin inhibitor
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1982 #sequence_revision 30-Jun-1987 #text_change 08-Dec-2000
C:Accession: S13433; S10778; A93642; A90225; A90686; PM0450; B39079; A61580;
3217
R:Vetr, H.; Gebhard, W.
Biol. Chem. Hoppe-Seyler 371, 1185-1196, 1990
A:Title: Structure of the human alpha(1)-microglobulin-bikunin gene.
A:Reference number: S13433; MUID:91214554
A:Accession: S13433
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-352 <VERT1>
A:Cross-references: EMBL:X54816; NID:924475; PIDN:CAA38585.1; PID:9825614; EMBL:X54817;
R:Diaria-Mehrpour, M.; Bourguignon, M.; Bourguignon, R.; Seshboue, R.; Sallier, J.P.; Leveillard, T.; Mart
Eur. J. Biochem. 191, 131-139, 1990
A:Title: Structural analysis of the human inter-alpha-trypsin inhibitor light-chain gene
A:Reference number: S10778; MUID:90336621
A:Accession: S10778
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <DIA>
R:Kaumeyer, J.F.; Polazzi, J.O.; Kotlick, M.P.
Nucleic Acids Res. 14, 7839-7850, 1986
A:Title: The mRNA for a proteinase inhibitor related to the HI-30 domain of inter-alpha-
A:Reference number: A93642; MUID:87040757
A:Accession: A93642
A:Molecule type: mRNA
A:Residues: 1-352 <KAU>
A:Cross-references: GB:X04494; NID:924478; PIDN:CAA28182.1; PID:924479
R:Lopez Olin, C.; Grubb, A.O.; Mendez, E.
Arch. Biochem. Biophys. 228, 544-554, 1984
A:Title: The complete amino acid sequence of human complex-forming glycoprotein heteroge
A:Reference number: A90074; MUID:84126849
A:Accession: A90074
A:Molecule type: protein
A:Residues: 20-56,58-202 <LOP>
A:Experimental source: individual with tubular proteinuria
A:Note: no evidence of sequence heterogeneity could be found, in spite of persistent het
R:Takagi, T.; Takagi, K.; Kawai, T.
Biochem. Biophys. Res. Commun. 98, 997-1001, 1981
A:Title: Complete amino acid sequence of human alpha-1-microglobulin.
A:Reference number: A90225; MUID:81184038
A:Accession: A90225
A:Molecule type: protein
A:Residues: 20-47,58-136,138-141,143-144,146-198 <TAK>
A:Experimental source: pooled urine of patients with tubular proteinuria
R:Reisinger, P.; Hochstrasser, K.; Albrecht, G.J.; Lempert, K.; Sallier, J.P.
Biochem. Biophys. Res. Commun. 96, 479-483, 1985
A:Title: Human inter-alpha-trypsin inhibitor: localization of the kunitz-type domains in
A:Reference number: A90686; MUID:85225968
A:Accession: A90686
A:Molecule type: protein
A:Residues: 206-290, 'VI', 293-342, 'E', 344-350 <REI>
R:Altman, F.; Lacour, B.; Strecker, G.; Parry, P.; Druce, T.; Daudon, M.
Biochem. Biophys. Res. Commun. 191, 1158-1165, 1993
A:Title: Molecular characteristics of uronic-acid-rich protein, a strong inhibitor of ca
A:Reference number: PN0450; MUID:93221481
A:Accession: PN0450
A:Molecule type: protein
A:Residues: 206-214, 'X' <ANM1>
R:Engblid, J.J.; Salvesen, G.; Hefta, S.A.; Thøgersen, I.B.; Rutherford, S.; Pizzo, S.V.
J. Biol. Chem. 266, 747-751, 1991
A:Title: Chondroitin 4-sulfate covalently cross-links the chains of the human blood prot
A:Reference number: A39079; MUID:91093267
A:Accession: B39079
A:Molecule type: protein
A:Residues: 206-225 <ENCI>
R:Chitt, F.; Balduyck, M.; Mizon, C.; Laroui, S.; Sautiere, P.; Mizon, J.
Int. J. Biochem. 23, 1201-1203, 1991
A:Title: A chondroitin-sulfate chain is located on serine-10 of the urinary trypsin inh
A:Reference number: A61580; MUID:92175157
A:Accession: A61580
A:Molecule type: protein
A:Residues: 214, 'X', 216-222, 'X' <CHI>
R:McKeenan, W.L.; Sakagami, Y.; Hoshi, H.; McKeenan, K.A.
J. Biol. Chem. 261, 5378-5383, 1986
A:Title: Two apparent human endothelial cell growth factors from human hepatoma cells ar
A:Reference number: A92583; MUID:86168278
A:Accession: B25604
A:Molecule type: protein
A:Residues: 206-214, 'X', 216-230, 'X', 232-239, 'X', 241-248, 'XX', 251-252, 'X', 254 <MCK>
R:Engblid, J.J.; Thøgersen, I.B.; Pizzo, S.V.; Salvesen, G.
J. Biol. Chem. 264, 15975-15981, 1989
A:Title: Analysis of inter-alpha-trypsin inhibitor and a novel trypsin inhibitor, pre-al

A:Reference number: A92736; MUID:89380192
A:Accession: C34245
A:Molecule type: protein
A:Residues: 206-225 <ENG2>
R:Raboni, C.; Cortese, R.
Nucleic Acids Res. 14, 6340, 1986
A:Title: Sequence of a full length cDNA coding for human protein HC (alpha-1-microglo
A:Reference number: A25303; MUID:86312901
A:Accession: A25303
A:Molecule type: mRNA
A:Residues: 1-218, 'HW' <TRA>
A:Note: this mRNA sequence appears to contain errors after residue 218
R:Calero, M.; Escobedo, J.; Grubb, A.; Mendez, E.
J. Biol. Chem. 269, 384-389, 1994
A:Title: Location of a novel type of interpolypeptide chain linkage in the human prot
A:Reference number: A53110; MUID:94103241
A:Accession: A53110
A:Molecule type: protein
A:Residues: 45-57 <CAL1>
R:Veltr, H.; Koeigler, M.; Gebhard, W.
FEBS Lett. 245, 137-140, 1989
A:Title: The domain structure of the inhibitor subunit of human inter-alpha-trypsin i
A:Reference number: S03552; MUID:89171290
A:Accession: S03552
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 206-352 <VEI2>
R:Malik, N.; Balduyck, M.; Maes, P.; Capon, C.; Mizon, C.; Han, K.K.; Tartar, A.; Fou
Biochem. Biophys. Res. Commun. 191, 1005-1018, 1992
A:Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolat
A:Reference number: S28928; MUID:93039735
A:Accession: S28928
A:Molecule type: protein
A:Residues: 206-215 <MAL>
R:Korjelle, W.; Capon, C.; Balduyck, M.; Sautiere, P.; Kouach, M.; Michalski, C.; Four
Eur. J. Biochem. 221, 881-888, 1994
A:Title: Chondroitin sulphate covalently cross-links the three polypeptide chains of
A:Reference number: S43466; MUID:94229087
A:Accession: S43466
A>Status: preliminary
A:Molecule type: protein
A:Residues: 206-221 <MOR>
R:Wisniewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilcek, J.
Biochemistry 33, 7423-7429, 1994
A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable co
A:Reference number: A53642; MUID:94271799
A:Accession: A53642
A>Status: preliminary
A:Molecule type: protein
A:Residues: 206-217 <WIS>
R:Calero, M.; Mendez, E.; Garcia, E.
Biochim. Biophys. Acta 1249, 91-99, 1995
A:Title: Expression of the human complex-forming glycoprotein HC (alpha-1-microglobul
A:Reference number: S55688; MUID:95284116
A:Accession: S55688
A:Molecule type: protein
A:Residues: 20-24 <CAL2>
R:Bourguignon, J.; Diaria-Mehrpour, M.; Seshboue, R.; Frain, M.; Sala-Trepat, J.M.; Ma
Biochem. Biophys. Res. Commun. 131, 1146-1153, 1985
A:Title: Human inter-alpha-trypsin-inhibitor: characterization and partial nucleotide
A:Reference number: I52208; MUID:86025577
A:Accession: I52208
A:Molecule type: protein
A:Residues: 302-352 <BOU>
A:Cross-references: GB:M11562; NID:9186587; PIDN:AAA59194.1; PID:9307077
R:Wojcik, E.G.C.; van den Berg, M.; van der Linden, I.K.; Poort, S.R.; Cupers, R.; Be
Biochem. J. 311, 753-759, 1995
A:Title: Factor IX zuphen: a Cys(18) -> Arg mutation results in formation of a heter
A:Reference number: S59509; MUID:96067589
A:Accession: S59509
A:Molecule type: protein

A:Residues: 27-35,'Y',37 <MOJ>
 R:Atmanl, F.; Mizon, J.; Khan, S.R.
 Eur. J. Biochem. 236, 984-990, 1996
 A:Title: Identification of uronic-acid-rich protein as urinary bikunin, the light chain
 A:Reference number: S66434; MUID:96270753
 A:Accession: S66434
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-230 <ATM2>
 R:Akerstrom, B.; Bratt, T.; Enghild, J.J.
 FEBS Lett. 362, 50-54, 1995
 A:Title: Formation of the alpha(1)-microglobulin chromophore in mammalian and insect cell
 A:Reference number: S68728; MUID:95212882
 A:Accession: S68728
 A:Molecule type: protein
 A:Residues: 89-100 <AKE>
 R:Jessen, T.E.; Faarvang, K.L.; Ploug, M.
 FEBS Lett. 230, 195-200, 1988
 A:Title: Carbohydrate as covalent crosslink in human inter-alpha-trypsin inhibitor: a no
 A:Reference number: S02431; MUID:88167187
 A:Accession: S02431
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-217 <JES>
 R:Lopez, C.; Grubb, A.; Mendez, E.
 FEBS Lett. 144, 349-353, 1982
 A:Title: Human protein HC displays variability in its carboxyl-terminal amino acid sequ
 A:Reference number: A91304
 A:Contents: annotation; variant of alpha-1-microglobulin
 A>Note: pooled urine samples contained two forms of this protein, both lacking 57-Lys an
 R:Hochstrasser, K.; Schonberger, O.L.; Rossmannlth, I.; Wächter, E.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 1357-1362, 1981
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-
 by affinity chromatography.
 A:Reference number: A91698; MUID:82074265
 A:Contents: annotation; carbohydrate binding sites
 R:Morill, M.; Travis, J.
 Biol. Chem. Hoppe-Seyler 366, 19-21, 1985
 A:Title: The reactive site of human inter-alpha-trypsin inhibitor is in the amino-termi
 A:Reference number: A90682; MUID:8522540
 A:Contents: annotation; inhibitory sites
 A>Note: In vitro, the first twelve residues of the amino end of the inhibitor appear to
 A:Comment: Alpha-1-microglobulin and inter-alpha-trypsin inhibitor are proteolytically
 C:Comment: Alpha-1-microglobulin occurs in many physiological fluids including plasma, u
 . It contains at least one brown-yellow chromophore.
 Query Match 7.5%; Score 213; DB 1; Length 352;
 Best Local Similarity 40.2%; Pred. No. 7.1e-06;
 Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;
 Oy 274 CTGSPRLVLMHYDROGCGMTFPARGCDGAARCFEYEAQQQACARPGDACLPAVQ 333
 Db 240 CMGMTSRIF---YNGTSNACETFOYGGCGMGNMFVTEKECLQCTRTVA--ACNLPYVR 293
 Oy 334 GPCRGMEPRMAYSPLLQOCHPFVYGGCEGNGNMFHSRESCEDACPVP 380
 Db 294 GPCRAFIQLMAFDAYKGVKCVLFPGGCGGNGKFFSEKRCREYCGVP 340
 RESULT 7
 723573
 hypochromic protein K10D3.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T23573
 R:McMurray, A.
 submitted to the EMBL Data Library, June 1996
 A:Accession: T23573
 A:Reference number: Z19762
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-922 <WII>
 A:Cross-references: EMBL:Z75545; PIDN:CAA99886.1; GSPDB:GN00019; CESP:K10D3.4
 A:Experimental source: clone K10D3

C:Genetics:
 A:Gene: CESP:K10D3.4
 A:Map position: 1
 A:Introns: 60/1; 278/1; 355/1; 743/1; 802/1; 885/2
 Query Match 7.4%; Score 211; DB 2; Length 922;
 Best Local Similarity 19.6%; Pred. No. 2.2e-05;
 Matches 108; Conservative 45; Mismatches 165; Indels 234; Gaps 24;
 Oy 1 CP-NQLSPMLMVD-----AQSTCERECSDOCAAEEKCINVCGLHSCVAARFG 50
 Db 111 CPLNGNNPFRIFINQYOQCYGADATCEALGALSTYC-----ICSDDCVTNN-PI 160
 Oy 51 SPAAPTTAASC-----EGFYVCPOQSDCDIDGQPV---CR-----CRDR 87
 Db 161 SECCPSRAFAACIQPNEGTYPPGCGTTLNHWYHDPTGECRELKYGQYGANNFQTKDH 220
 Oy 88 CEKEPSFTCASDGLTYNN-----RCVYDAEACLRGLHLHYPCKHVLSWPPSSP 136
 Db 221 CESYCKQTC-NRGLPLYRDRRTTGKQEPVYCGQDNGC-----NNP 260
 Oy 137 GPPEETARPTPGAP-----VPPALYSPSPQAVQVGTASLHCDV-----SGR 180
 Db 261 NQCCTIMGLTQCCPFYLFICSRNGGIPSEVYNT-----AGLPTEYDVQIIPGSGN 313
 Oy 181 PPAYTWEKQSHORENLIMRPMQYNNVVTSTIGLVLYNARPEADAGLYCTARNAAGL 240
 Db 314 TSFRFYDSNE-----GRCIQFS-----YLQGGNFMNFI 343
 Oy 241 RADPFLSYVQR-----EPARDA-----APSPAPARC----- 267
 Db 344 SQDHCKEFCRILCSAGEPLKDSGERNMECSPTSGANSCTHSCESTSGTTGGVC 403
 Oy 268 -----LPDVOACTGPTSPHLVLMHYDPORGCGMTFPARGCDGAARCFEYEAQQA 318
 Db 404 CPRPQYVKLPARGNGCGTYSNR---WVFNAKGNCEEFYISCGGANNFETKEDQDY 460
 Oy 319 C--ARG-----PDA--CV 328
 Db 461 CRDAREPQCIQGTALTDNSGNFIICGSSAATSTCPANHYCYDDTTYCCPTQAVTCS 520
 Oy 329 LPVQVG--PORGMEPRMAYSPLLQOCHPFVYGGCEGNGNMFHSRESCEDACPVPRTF--- 383
 Db 521 LSTKSGASCPAVTRMYVYSTTRTCOTYSFNGGDSNMFATQDDCKYCRVESCPDGE 580
 Oy 384 -----PCRAC 388
 Db 581 VMKEQNGAARAC 592
 RESULT 8
 alpha-1-microglobulin / inter-alpha-trypsin inhibitor precursor [validated] - bovine
 N:Alternate names: BI-14 (inhibitory fragment of ITI); bikunin; ITI
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 25-Feb-1985 #sequence_revision 04-Feb-2000 #text_change 18-Aug-2000
 C:Accession: S68149; A91717; A90685; S31219; A01209
 R:Lindqvist, A.; Akerstrom, B.
 Biochim. Biophys. Acta 1306, 98-106, 1996
 A:Title: Bovine alpha(1)-microglobulin/bikunin. Isolation and characterization of 11v
 A:Reference number: S68149; MUID:96201710
 A:Accession: S68149
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-352 <LINA>
 A:Cross-references: EMBL:U35642; NID:g1016297; PIDN:AAB07599.1; PID:g1016298
 R:Hochstrasser, K.; Wächter, E.
 Hoppe-Seyler's Z. Physiol. Chem. 364, 1679-1687, 1983
 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte
 A:Reference number: A91717; MUID:84133807
 A:Accession: A91717
 A:Molecule type: protein

[illegible]

	Query Match	7.1%	Score 203;	DB 2;	Length 349;	
	Best Local Similarity	40.6%;	Pred. No. 2.8e-05;			
	Matches 39;	Conservative 13;	Mismatches 42;	Indels 2;	Gaps 1	
QY	285 MHYPDQRGGCTPFARGCDGAARGFETYEACOAARGRBDCVLPAYVGPCRGMEPRWA	344				
	:::::	:	: : : :	:	:	
Dd	246 YYYGASACETFQYGGCLGNNGNNFTISEKDLQTCTRTI--ACNLPYVGPCRAFIKTMLA	303				
QY	345 YSPLLDQCCHPEVVYGCGEENGNNFHSRESCDCAPR	380				
	:::	:	: : : :	:	:	
Dd	304 FDAAAGCKTDFHYGGCKKGNGNKFYSEKECKEYCVP	339				

```

RESULT      14
T26859
hypothetical protein Y43F8B.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26859
R:Ainscough, R.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20278
A:Accession: T26859
A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 1-1743 <WILL>
A:Cross-references: EMBL:AL032623; PIDN:CAA21511.1; CESP:Y43F8B.3
A:Experimental source: clone Y43F8B
C:Genetics
A:Gene: CESP:Y43F8B.3
A:Introns: 65/1; 92/2; 128/1; 229/1; 367/1; 422/1; 486/2; 523/1; 571/1; 628/1; 857/2;

```

Query Match 7.1%; Score 202.5; DB 2; Length 1743;
 Best Local Similarity: 22.1%; Pred. No. 0.0013;
 Matches 124; Conservative 49; Mismatches 235; Indels 153; Gaps 26;

QY 20 ECSRDDCAAAREKCCINNVGLHSCVAARFPGSPAAP-----TTAASCE 62
 Db 1019 QCSPOQVCGPGHYCHVGEANYCCKA--LGSDPCGQPLDRGVGSGQLSRMYWNOQSQCCL 1076
 QY 63 GF-VCPQOGS-----DCIINDGQPCRCRDCRCEKSPSTCASDGLTYNNR---CYMD 110
 Db 1077 PFSYCGQGTQNNFLTYKODC-----RTCYELDNPALGDPOQAQNNRPLQCSAT 1126
 QY 111 AEACLRGLHLHIVPCKHVLSPSSPCPPTTARPT---PPAAPVPALYSSPQAVQV 167
 Db 1127 ASTCGAGFWCHFGANQDTTCCPGRVESPOLCOQPMANGTGATLPRRYVYNAQTQCCQV 1186
 QY 168 GGTASLHCDVSGRPRAVYTWKQSHQRENILMRPDQYGNVYVTSIGOLVLYNARPEADAG 227
 Db 1187 -----NYAGR-----MGQNNFQSQACEQTCVPYVYVNCPTGSPMLDASTNKP----- 1229
 QY 228 LYTGCT-ARNAAGLRADF-PLSYVQRE-PARDAAPSTPAEACLPDVOACTGPTSPHLY 283
 Db 1230 -VPTEFGSNSCG---ADHCHLGLVPDEYQCCPSSPTNPGACQGLPESEGVYGAAPAPPTS 1285
 QY 284 LMHYDPQRCGCMTPPARCGDGAARGFETYEACQOAC-----ARGPDA 326
 Db 1286 RMYTDQDMQCKQFTYNGRBNQNNFLQEDCATCDVFTNPNQPIALPATICSGTSS 1345
 QY 327 -----CVLPAYQCP---CRGNEPRWVYSPLLQCHPEFYGGCEGNGNPFHSESCED 375
 Db 1346 DTCCANMWHIGANQDSTVCC---PSGRKSKNFQOCLPQYNGRGNQNNFENQADDER 1401
 QY 376 ACPVPRP-----PCRACRLRSKLTSLCRS--DEFAVGRLE----- 411
 Db 1402 TCPPIFNCLGEVILIEDCAPRCK-----PLKNSCGSAEFCHTGPSONSFCPCR 1455
 QY 412 VLEBP-----BAAGGIARVALEDVLDK---DKMGLK-----FLGTXYLEVTLSCMD 454
 Db 1456 INODPCNAFVNGSGNFMRTYRYPVVEGDGFSFOYRGLKGNENNFLTKMQQETCRPLT 1515
 QY 455 WACPCPNMTAGDPLVYINGEV 475
 Db 1516 TAC-----FGGESPLMNGRV 1531

RESULT 15
 S28291
 hypothetical protein ZC84.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Aug-1997
 C:Accession: S28291
 R:Thomas, K.
 Submitted to the EMBL Data Library, December 1992
 A:Reference number: S28285
 A:Accession: S28291
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2844 <THO>
 A:Cross-references: EMBL:Z19157
 C:Genetics:
 A:Introns: 14/1; 32/3; 57/1; 192/3; 277/1; 398/1; 439/1; 474/1; 497/1; 813/1; 1135/1; 12
 493/1; 2555/1; 2720/1; 2739/3; 2819/1
 C:Superfamily: animal Kunitz-type proteinase inhibitor homology
 F:220-274/Domain: animal Kunitz-type proteinase inhibitor homology <BP11>
 F:343-395/Domain: animal Kunitz-type proteinase inhibitor homology <BP12>
 F:442-492/Domain: animal Kunitz-type proteinase inhibitor homology <BP13>
 F:546-598/Domain: animal Kunitz-type proteinase inhibitor homology <BP14>
 F:654-706/Domain: animal Kunitz-type proteinase inhibitor homology <BP15>
 F:1662-1716/Domain: animal Kunitz-type proteinase inhibitor homology <BP16>
 F:1787-1839/Domain: animal Kunitz-type proteinase inhibitor homology <BP17>
 F:1845-1895/Domain: animal Kunitz-type proteinase inhibitor homology <BP18>
 F:1952-2004/Domain: animal Kunitz-type proteinase inhibitor homology <BP19>
 F:2097-2152/Domain: animal Kunitz-type proteinase inhibitor homology <BP10>

Query Match 7.0%; Score 200.5; DB 2; Length 2844;
 Best Local Similarity: 21.9%; Pred. No. 0.00026;
 Matches 105; Conservative 47; Mismatches 175; Indels 153; Gaps 20;

QY 21 CSRDDCAAAREKCCINNVGLHSCVAARFPGSPAAPTTAASCEGFYCPQOQSDCDIMDGP 80
 Db 1470 CAFNTDCLAGFNCNGLC---SCL-----TTIAABEGY-CYK-----IDPQGP 1509
 QY 81 VCRCRDCEKE-PSFTCASD-----GLTYNCTYDAEACLRGLHLH-----IVPKH 127
 Db 1510 GCVYNEQCAVWPDAFCDTSAVGCTCRCEKKEKVERATRDGHVCLVDLDAHNNTLAITC-- 1567
 QY 128 VLSWPPSSPCPPTTA-----RTPGAAPVPALYSSPQAVQVYG----- 169
 Db 1568 -----PLDEGAGVYSALSDPKHPRONDSPGVLCTNTDSTNTNQGDQAGDGSACLPFSN 1622
 QY 170 -----TASLHCDVSGRPRAVYTWKQSHQRENILMRPDQYGNVYVTSIG-- 214
 Db 1623 GQIYADQYDCEYFVSSLDLTSSG-----YSEKANGICCPNRAFTCVPTATGPN 1671
 QY 215 -----QLVLYNARPEADGLY-----TCTARNAAGLIRA 242
 Db 1672 PTEPRMYNYSITGMCOQFLMPDPSASGSEHSPNNFRVYEHCEFCRDTCSGAPETIHR 1731
 QY 243 DF-----PLSYVQREDA---RDAAPSIAPABCLPDVOACTGPT-----SPHL-----VLMH 286
 Db 1732 SFLEQTPITGCSQYSACSHNTECKSVSTQWCPCPVASVCGPVGSRPLDPSVNNNGTYIH 1791
 QY 287 -----YDPRGCGMTFPARGCDGAARGFETYEACQOACARPGDQCVLPA 331
 Db 1792 SGVEKQGTATSTRFPYDPOGSKCSFTYLGAGGNNNFLSRIDELIYCAR--XTICTOPL 1849
 QY 332 VQGPCRGNEPRWVYSPLLQCHPEFYGGCEGNGNPFHRECEDEAC---VYPRTPPCRA 387
 Db 1850 RVGNCDRSVRRYVSAATRECOSEFYTCQGDNDNFELVDQYQFCRNAAPPRCPQOGA 1909

Search completed: February 26, 2002, 01:29:49
 Job time: 428 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:28:07 ; Search time 69.26 Seconds

(without alignments)
167.654 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_548

Perfect score: 2855
Sequence: 1 CPNGLSPNLMVADASTCERE.....KKILELLEKQACELNRFQD 516Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	7.5	143	US-08-422-333-10	Sequence 10, Appl
2	213	7.5	143	5223482-20	Patent No. 5223482
3	213	7.5	144	5187153-18	Patent No. 5187153
4	213	7.5	147	US-08-358-160-72	Sequence 72, Appl
5	210	7.4	123	5466783-22	Patent No. 5466783
6	208	7.3	123	5466783-21	Patent No. 5466783
7	207	7.3	122	US-08-422-333-12	Sequence 12, Appl
8	207	7.3	122	5187153-20	Patent No. 5187153
9	207	7.3	122	5220013-23	Patent No. 5220013
10	207	7.3	127	5466783-24	Patent No. 5466783
11	206	7.2	122	5223482-22	Patent No. 5223482
12	200.5	6.9	122	5466783-23	Patent No. 5466783
13	196	6.9	123	5466783-25	Patent No. 5466783
14	194.5	6.8	276	US-07-828-920A-1	Sequence 1, Appl
15	194.5	6.8	276	US-08-437-841-9	Sequence 9, Appl
16	194.5	6.8	276	US-08-286-521-9	Sequence 9, Appl
17	194.5	6.8	276	US-08-436-175-9	Sequence 9, Appl
18	194.5	6.8	276	US-08-796-850-1	Sequence 1, Appl
19	194.5	6.8	276	US-08-854-764-3	Sequence 3, Appl
20	194.5	6.8	276	US-08-943-682-9	Sequence 9, Appl
21	194.5	6.8	276	PCR-US95-09377-3	Sequence 3, Appl
22	194.5	6.8	276	PCR-US95-09464-9	Sequence 9, Appl
23	194.5	6.8	277	US-07-844-297-1	Sequence 1, Appl
24	194.5	6.8	304	US-08-026-145-2	Sequence 2, Appl
25	194.5	6.8	304	US-08-446-646-9	Sequence 9, Appl
26	194.5	6.8	304	US-08-676-125A-18	Sequence 18, Appl
27	194.5	6.8	304	US-09-136-012A-18	Sequence 18, Appl

28	194.5	6.8	304	US-08-676-124-1	Sequence 1, Appl
29	194.5	6.8	304	US-08-208-264A-25	Sequence 25, Appl
30	194.5	6.8	304	US-09-414-878-1	Sequence 1, Appl
31	194.5	6.8	304	US-09-240-136-1	Sequence 1, Appl
32	194.5	6.8	304	US-09-054-782-2	Sequence 2, Appl
33	194.5	6.8	304	5466783-2	Patent No. 5466783
34	194.5	6.8	352	US-08-854-764-2	Sequence 2, Appl
35	194.5	6.8	352	PCR-US95-09377-2	Sequence 2, Appl
36	192	6.7	252	US-08-685-660A-7	Sequence 7, Appl
37	192	6.7	252	US-08-974-196-7	Sequence 7, Appl
38	192	6.7	252	US-09-071-709-10	Sequence 10, Appl
39	192	6.7	252	US-09-013-896A-2	Sequence 2, Appl
40	179.5	6.3	235	US-08-147-710-2	Sequence 2, Appl
41	179.5	6.3	235	US-08-458-090-2	Sequence 2, Appl
42	179.5	6.3	235	US-08-457-887-2	Sequence 2, Appl
43	179.5	6.3	235	US-08-817-145-3	Sequence 3, Appl
44	179	6.3	213	US-08-796-850-2.	Sequence 2, Appl
45	178	6.2	83	5220013-22	Patent No. 5220013

ALIGNMENTS

```
RESULT 1
US-08-422-333-10
; Sequence 10, Application US/08422333
; Patent No. 5912410
; GENERAL INFORMATION:
; APPLICANT: CORDELL, Barbara L.
; TITLE OF INVENTION: TRANSGENIC NON-HUMAN MAMMAL DISPLAYING
; TITLE OF INVENTION: THE AMYLOID-FORMING PATHOLOGY OF ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scios, Inc.
; STREET: 2450 Bayshore Parkway
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,333
; FILING DATE: 13-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Shearer, Peter R.
; REGISTRATION NUMBER: 28,117
; REFERENCE/DOCKET NUMBER: 21900-28048.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 966-1550
; TELEFAX: (415) 968-2438
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 143 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-422-333-10

Query Match 7.5%; Score 213; DB 2; Length 143;
Best Local Similarity 39.3%; Pred. No. 8.3e-08;
Matches 42; Conservative 13; Mismatches 46; Indels 6; Gaps 2;

QY 274 CIGCTSPHVLVWVHYDQRCGCTFPARGCDGARGETYEACQACARCGDAGCTVPANO 333
DB 35 CIMGTRSYR-----YNGTSMACETFOYGGCGMGNNNVTEKECIQTCRTVA--ACNIPVIR 88
QY 334 GPCRGWEPHMAVSPILQDCHPFIYGGCEGNGNHFHRESEEDACVPV 380
```

Db 89 GPCRAFIQLMAFDVAKGCVLPYGGCGNGNKFYSEKREYCGVP 135

RESULT 2
5223482-20

Patent No. 5223482
APPLICANT: SCHILLING, JAMES W.; PONTE, PHYLLIS A.; CORDELL,
BARBARA
TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
NUMBER OF SEQUENCES: 34

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/361,912

FILING DATE: 06-JUN-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 359,911

FILING DATE: 12-MAY-1989

APPLICATION NUMBER: 87,002

FILING DATE: 18-AUG-1987

APPLICATION NUMBER: 8,810

FILING DATE: 30-JAN-1987

APPLICATION NUMBER: 948,376

FILING DATE: 31-DEC-1986

APPLICATION NUMBER: 932,193

FILING DATE: 17-NOV-1986

SEQ ID NO: 20:

LENGTH: 143

5223482-20

Query Match 7.5%; Score 213; DB 6; Length 143;

Best Local Similarity 39.3%; Pred. No. 8.3e-08;

Matches 42; Conservative 13; Mismatches 46; Indels 6; Gaps 2;

QY 274 CTGPTSPHLVLMHYDPQGGCTFPARGCDGARGFEYEAQACAGPDACVLPVQ 333

Db 35 CMGMSRFF---YNGTSMACETFOYGGCMGNGNMFVEKECLQCRIVA--ACNLPIVR 88

QY 334 GPCRGMEPRMAYSPILQOCHPFFVYGGCEGNGNMFHRSRSCDACPVP 380

Db 89 GPCRAFIQLMAFDVAKGCVLPYGGCGNGNKFYSEKREYCGVP 135

RESULT 3
5187153-18

Patent No. 5187153

APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S
AMYLOID POLYPEPTIDE DERIVATIVES
NUMBER OF SEQUENCES: 33

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/502,273

FILING DATE: 29-MAR-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 361,912

FILING DATE: 06-JUN-1989

APPLICATION NUMBER: 359,911

FILING DATE: 12-MAY-1989

APPLICATION NUMBER: 87,002

FILING DATE: 18-AUG-1987

APPLICATION NUMBER: 8,810

FILING DATE: 30-JAN-1987

APPLICATION NUMBER: 948,376

FILING DATE: 31-DEC-1986

APPLICATION NUMBER: 932,193

FILING DATE: 17-NOV-1986

SEQ ID NO: 18:

LENGTH: 144

5187153-18

Query Match 7.5%; Score 213; DB 6; Length 144;

Best Local Similarity 39.3%; Pred. No. 8.4e-08;
Matches 42; Conservative 13; Mismatches 46; Indels 6; Gaps 2;

QY 274 CTGPTSPHLVLMHYDPQGGCTFPARGCDGARGFEYEAQACAGPDACVLPVQ 333

Db 35 CMGMSRFF---YNGTSMACETFOYGGCMGNGNMFVEKECLQCRIVA--ACNLPIVR 88

QY 334 GPCRGMEPRMAYSPILQOCHPFFVYGGCEGNGNMFHRSRSCDACPVP 380

Db 89 GPCRAFIQLMAFDVAKGCVLPYGGCGNGNKFYSEKREYCGVP 135

RESULT 4

US-08-358-160-72

Sequence 72 Application US/08358160

Patent No. 5663143

GENERAL INFORMATION:

APPLICANT: LEY, Arthur C.

APPLICANT: LADNER, Robert C.

APPLICANT: GUTERMAN, Sonia K.

APPLICANT: ROBERTS, Bruce L.

APPLICANT: MARKLAND, William

APPLICANT: KENT, Rachel B.

TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ

NUMBER OF SEQUENCES: 234

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W. Suite 300

CITY: Washington

STATE: District of Columbia

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/358,160

FILING DATE: 16-DEC-1994

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/133,031

FILING DATE: 13-OCT-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/009,319

FILING DATE: 26-JAN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/664,989

FILING DATE: 01-MAR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/487,063

FILING DATE: 02-MAR-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/240,160

FILING DATE: 02-SEP-1988

ATTORNEY/AGENT INFORMATION:

NAME: Cooper, Iver P.

REGISTRATION NUMBER: 28,005

REFERENCE/DOCKET NUMBER: LEY-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 147 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-358-160-72

Query Match 7.5%; Score 213; DB 1; Length 147;
Best Local Similarity 40.2%; Pred. No. 8.6e-08;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

OY 274 GPTSPHLVLMHYDPORGCGMTFPARGCDGAARGFETYEACQACARGPDACVLPAYOGP 333
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 35 CGMTSRRF---YNGTSMACETFOYGGCMGNGNMFVTEKECLQTCRYA--ACNLPYV 88
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 334 GRCGMEPRMAYSPLLQOCHPFVYGGCEGNGNMFHSRSCEDACVP 380
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 89 GCRAFIOLMAFDVAVKCVLFYGGCGGNGNKFYSKCKEYCGVP 135

RESULT 5
546783-22
; Patent No. 546783
; APPLICANT: Mun, Tze-Cheln.; Kretzmer, Kuniko K.; Broze,
; George J. Jr.
; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/93,285
; FILING DATE: 15-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 566,280
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 123,753
; FILING DATE: 23-NOV-1987
; APPLICATION NUMBER: 77,366
; FILING DATE: 23-JUL-1987
; SEQ ID NO:22:
; LENGTH: 123
546783-22

Query Match 7.4%; Score 210; DB 6; Length 123;
Best Local Similarity 39.0%; Pred. No. 1.1e-07;
Matches 41; Conservative 14; Mismatches 48; Indels 2; Gaps 1;
OY 276 GPTSPHLVLMHYDPORGCGMTFPARGCDGAARGFETYEACQACARGPDACVLPAYOGP 335
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 12 GPCLGMIKRYFYNGTSMACETFOYGGCMGNGNMFVTEKECLQTCR--TVSACSLPIYOGP 69
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 336 GRCGMEPRMAYSPLLQOCHPFVYGGCEGNGNMFHSRSCEDACVP 380
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 70 CRAFTILMAFDVAQKCVLFYGGCGGNGNKFYSKCKEYCGVP 114

RESULT 6
546783-21
; Patent No. 546783
; APPLICANT: Mun, Tze-Cheln.; Kretzmer, Kuniko K.; Broze,
; George J. Jr.
; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/93,285
; FILING DATE: 15-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 566,280
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 123,753
; FILING DATE: 23-NOV-1987
; APPLICATION NUMBER: 77,366
; FILING DATE: 23-JUL-1987
; SEQ ID NO:21:
; LENGTH: 123
546783-21

Query Match 7.3%; Score 208; DB 6; Length 123;

Best Local Similarity 38.1%; Pred. No. 1.5e-07;
Matches 40; Conservative 14; Mismatches 49; Indels 2; Gaps 1;

OY 276 GPTSPHLVLMHYDPORGCGMTFPARGCDGAARGFETYEACQACARGPDACVLPAYOGP 335
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 12 GPCLGMIKRYFYNGTSMACETFOYGGCMGNGNMFVTEKECLQTCRYA--ACNLPYV 69
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 336 GRCGMEPRMAYSPLLQOCHPFVYGGCEGNGNMFHSRSCEDACVP 380
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 70 CRAFTILMAFDVAQKCVLFYGGCGGNGNKFYSKCKEYCGVP 114

RESULT 7
US-08-422-333-12
; Sequence 12, Application US/08422333
; Patent No. 5912410
; GENERAL INFORMATION:
; APPLICANT: CORDELL, Barbara L.
; TITLE OF INVENTION: TRANSGENIC NON-HUMAN MAMMAL DISPLAYING
; TITLE OF INVENTION: THE AMYLOID-FORMING PATHOLOGY OF ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scios, Inc.
; STREET: 2450 Bayshore Parkway
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,333
; FILING DATE: 13-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Shearer, Peter R.
; REGISTRATION NUMBER: 28,117
; REFERENCE/DOCKET NUMBER: 21900-28048.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 966-1550
; TELEFAX: (415) 968-2438
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 122 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-422-333-12

Query Match 7.3%; Score 207; DB 2; Length 122;
Best Local Similarity 38.1%; Pred. No. 1.8e-07;
Matches 40; Conservative 16; Mismatches 47; Indels 2; Gaps 1;
OY 276 GPTSPHLVLMHYDPORGCGMTFPARGCDGAARGFETYEACQACARGPDACVLPAYOGP 335
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 12 GPCLGMIKRYFYNGTSMACETFOYGGCMGNGNMFVTEKECLQTCR--TVSACSLPIYOGP 69
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
OY 336 GRCGMEPRMAYSPLLQOCHPFVYGGCEGNGNMFHSRSCEDACVP 380
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 70 CRAFTILMAFDVAVKCVLFYGGCGGNGNKFYSKCKEYCGVP 114

RESULT 8
5187153-20
; Patent No. 5187153
; APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
; TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S
; AMYLOID POLYPEPTIDE DERIVATIVES
; NUMBER OF SEQUENCES: 33

;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/502,273
;; FILING DATE: 29-MAR-1990
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 361,912
;; FILING DATE: 06-JUN-1989
;; APPLICATION NUMBER: 359,911
;; FILING DATE: 12-MAY-1989
;; APPLICATION NUMBER: 87,002
;; FILING DATE: 18-AUG-1987
;; APPLICATION NUMBER: 8,810
;; FILING DATE: 30-JAN-1987
;; APPLICATION NUMBER: 948,376
;; FILING DATE: 31-DEC-1986
;; APPLICATION NUMBER: 932,193
;; FILING DATE: 17-NOV-1986
;; SEQ ID NO:20:
;; LENGTH: 122
5187153-20

Query Match 7.3%; Score 207; DB 6; Length 122;
Best Local Similarity 38.1%; Pred. No.1.8e-07;
Matches 40; Conservative 16; Mismatches 47; Indels 2; Gaps 1;
QY 276 GPTSPHLVLMHYDPORGCMTPFARGCDGAARGFETYEACQACARGPDCACVLPAYQ 335
DB 12 GPCGLKFRFYNGTSMACETFLYGGCMGNLNNFLSQKECLQTCR--TVEACNLPYVGP 69
QY 336 CRGMEPRMAYSPILQOCHPFYVYGGCEGNGNNFHSRESCDACPVP 380
DB 70 CRAFTIQLMAFDVAVKGCYVRFSGCKGNKMFYSQKECKEYCGIP 114

RESULT 9
5220013-23
; Patent No. 5220013
; APPLICANT: PONTE, PHYLLIS A.;CORDELL, BARBARA
; TITLE OF INVENTION: DNA SEQUENCE USEFUL FOR THE DETECTION
; OF ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/444,118
; FILING DATE: 30-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO:23:
; LENGTH: 122
5220013-23

Query Match 7.3%; Score 207; DB 6; Length 122;
Best Local Similarity 38.1%; Pred. No.1.8e-07;
Matches 40; Conservative 16; Mismatches 47; Indels 2; Gaps 1;
QY 276 GPTSPHLVLMHYDPORGCMTPFARGCDGAARGFETYEACQACARGPDCACVLPAYQ 335
DB 12 GPCGLKFRFYNGTSMACETFLYGGCMGNLNNFLSQKECLQTCR--TVEACNLPYVGP 69
QY 336 CRGMEPRMAYSPILQOCHPFYVYGGCEGNGNNFHSRESCDACPVP 380
DB 70 CRAFTIQLMAFDVAVKGCYVRFSGCKGNKMFYSQKECKEYCGIP 114

RESULT 10
5466783-24

; Patent No. 5466783
; APPLICANT: Mun, Tze-Chen.;Kretzmer, Kuniko K.;Broze,
; George J. Jr.
; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/93,285
; FILING DATE: 15-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 566,280
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 123,753
; FILING DATE: 23-NOV-1987
; APPLICATION NUMBER: 77,366
; FILING DATE: 23-JUL-1987
; SEQ ID NO:24:
; LENGTH: 127
5466783-24

Query Match 7.3%; Score 207; DB 6; Length 127;
Best Local Similarity 40.4%; Pred. No.1.9e-07;
Matches 44; Conservative 12; Mismatches 45; Indels 8; Gaps 3;
QY 274 CTGTPSPHLVLMHYDPORGCMTPFARGCDGAARGFETYEACQACARGPDCACVLPAYQ 333
DB 14 CMGMTSRFY---YNGTSMACETFOYGGCMGNNGNNFTEKKECLQTCRVA--ACNLPYV 67
QY 334 GPCGMEPRMAYSPILQOCHPFYVYGGCEGNGNNFHSRESC--DACPVP 380
DB 68 GPCRAFTIQLMAFDVAVKGCYVLPYGGCGGNGNKFYSRKECKEYCGIP 116

RESULT 11
5223482-22
; Patent No. 5223482
; APPLICANT: SCHILLING, JAMES W.;PONTE, PHYLLIS A.;CORDELL,
; BARBARA
; TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
; INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/361,912
; FILING DATE: 06-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 359,911
; FILING DATE: 12-MAY-1989
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO:22:
; LENGTH: 122
5223482-22

Query Match 7.2%; Score 206; DB 6; Length 122;
Best Local Similarity 39.6%; Pred. No.2.1e-07;
Matches 38; Conservative 16; Mismatches 40; Indels 2; Gaps 1;
QY 285 WHDPORGCMTPFARGCDGAARGFETYEACQACARGPDCACVLPAYQPCRGMEPRMA 344
DB 23 YFNGTSMACETFLYGGCMGNLNNFLSQKECLQTCR--TVEACNLPYVGP CRAFTIQLMA 80
QY 345 YSPILQOCHPFYVYGGCEGNGNNFHSRESCDACPVP 380
DB 81 FDAVKGKCVRFSGCKGNKMFYSQKECKEYCGIP 116

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:26:45 ; Search time 162.6 Seconds

(without alignments)
235.067 Million cell updates/sec

Title: US-09-819-136-2_COPY_33_548

Perfect score: 2855
Sequence: 1 CPNGLSPNLMVADASTCERE.....KRIELLEKACELNRFQD 516

Scoring table:
BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SID2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	679	23.8	216	22	AA88333
2	358	12.5	59	21	AA84381
3	222.5	7.8	560	15	AA62523
4	213	7.5	143	20	AA08609
5	213	7.5	147	17	AA92237
6	213	7.5	147	18	AA25928
7	213	7.5	352	9	AA81110
8	213	7.5	352	22	AA014343
9	213	7.5	352	22	AA014579
10	213	7.5	352	22	AA05095
11	213	7.5	366	21	AA843730

12	212	7.4	147	19	AA69522
13	210	7.4	124	18	AA25932
14	210	7.4	144	18	AA25936
15	210	7.4	145	18	AA25935
16	210	7.4	165	18	AA25937
17	209	7.3	124	17	AA92233
18	209	7.3	124	18	AA25933
19	209	7.3	144	18	AA25938
20	209	7.3	145	17	AA92235
21	207	7.3	122	20	AA08611
22	195.5	6.8	122	13	AA824110
23	195.5	6.8	304	16	AA87389
24	194.5	6.8	256	12	AA811169
25	194.5	6.8	262	12	AA811172
26	194.5	6.8	262	12	AA811172
27	194.5	6.8	265	12	AA811170
28	194.5	6.8	276	17	AA892265
29	194.5	6.8	276	17	AA892012
30	194.5	6.8	276	18	AA830311
31	194.5	6.8	276	19	AA651535
32	194.5	6.8	304	10	AA892002
33	194.5	6.8	304	16	AA81884
34	194.5	6.8	304	16	AA867994
35	194.5	6.8	304	20	AA49557
36	194.5	6.8	304	21	AA870273
37	194.5	6.8	352	17	AA892011
38	192.5	6.7	277	14	AA837312
39	192	6.7	170	18	AA830041
40	192	6.7	170	21	AA814189
41	192	6.7	170	21	AA814190
42	192	6.7	179	18	AA830053
43	192	6.7	179	21	AA814159
44	192	6.7	197	18	AA830043
45	192	6.7	197	21	AA814160

ALIGNMENTS

RESULT 1	AA88333	standard; Protein; 216 AA.
ID	AA88333	
AC	AA88333	
DT	23-MAY-2001	(first entry)
DE	Human membrane or secretory protein clone pSEC0040.	
KW	Human; secretory protein; membrane protein; vaccine; gene therapy; rheumatoid arthritis; diabetes.	
OS	Homo sapiens.	
PN	EP1067182-A2.	
PD	10-JAN-2001.	
PF	07-JUL-2000; 2000EP-0114090.	
PR	08-JUL-1999; 99JP-0194179.	
PR	11-JAN-2000; 2000JP-0118775.	
PR	02-MAY-2000; 2000JP-0183766.	
PA	(HELI-) HELIX RES INST.	
PI	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;	
DR	WPI; 2001-093989/11.	
DR	N-PSDB; AAF93760.	
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -	

rUTI protein SEQ I
New protease inh
SUC2-Bp1-d21-RPDE
Novel protease inh
Elastase inhibitor
Human mature urina
New protease inh
SUC2-Bp7-d21-RPDE
Human urinary tryp
Inter-alpha-trypsi
Trypsin-like structu
Human lipoprotein-
Ser(Asp1-Thr255)-E
Ser(Asp1-Glu245)-
Ser(Asp1-Ser248)-
TfPI mutain, Lys36
Recombinant non-gl
Human tissue facto
Human tissue facto
Lipoprotein-associ
Tissue factor path
Human lipoprotein
Human tissue facto
Ubiquitin-TfPI fus
Non-glycosylated T
Human placental bl
Human placental bl
Human placental bl
Human placental bl
Mature human place
Human placental bl

SQ Sequence 216 AA;

QY 1 CPNGLSPNLWVDAQSTICERECSDQDDCAAEKCCINWCGLHCVAAKPPGSPAAPTTAAS 600
|||||
33 CPNGLSPNLWVDAQSTICERECSDQDDCAAEKCCINWCGLHCVAAKPPGSPAAPTTAAS 920
Db

9y CEGFVCPDGGSDCDIDWDGQPVCRRCRDREKEPSSFTCASDGLTYYNRCYMDAEACRLGHLT 120
61
|||
93 ceafvcpdaasdcddwqamuvrcrdrfakaneftasdcrltutuevmudaoacccca-- 140
|||

QY 121 HTPCKHVLWSPSP--SPGPEETARTPTGGAAPVPAALYSSSPQAVQVGGTASLHCDVS 178

150 ----ctst-scptscsagtpaargrrplpap-----hlgrpr-----cllpctaa 191

```

QY      179 GRPPPAVTWEKOSHORENLIMRP 201
          ||          |:| | :||
Db      192 ---pp-----hrrcrlyvr 203

```

RESULT	2
AAV84381	
ID	AAV84381 standard; Protein; 59 AA

AC AAY84381;

DT 12-JUL-2000 (first entry)

Amino acid sequence of a serine proteinase inhibitor designated ZKUN6.

serine proteinase inhibitor; knitt1 domain; ZKRN6; acute pancreatitis;
cardiopulmonary bypass-induced pulmonary injury; myocardial infarction;
allergy-induced proteinase release; deep vein thrombosis; shock;
hyperfibrinolytic haemorrhage; emphysema; rheumatoid arthritis;
adult respiratory distress syndrome; chronic inflammatory bowel disease;
psoriasis; inflammatory conditions; platelet function; fibrinolysis;
organ preservation; wound healing; haemostasis imbalance; cirrhosis;
acquired coagulopathy; primary fibrinolysis; thrombolytic therapy;
blood coagulation; proteolytic tissue degradation; gene therapy.

05 Homo sapiens

XX	Key	Location/Qualifiers
----	-----	---------------------

FT	Disulfide-bond	6.56
FT	Domain	14.20

FT	Disulfide-bond	15..39
FT	Disulfide-bond	31..52

PN WO200014235-A1

PD 16-MAR-2000.

PF 01-SEP-1999; 99WO-US20202.

PR 03-SEP-1998; 98US-0148092.
yy

PA (ZYMO) ZYMOGENETICS INC.
YY

Conklin DC;
PI XY

DR WPI; 2000-256985/22.
DR N-PSDB: AA700003 AA700003

XX	XX	Novel	proteinase inhibitor
DT			

PT Novel proteinase inhibitor zkun6 comprising a kunitz domain, useful in the treatment or prevention of conditions associated with excessive proteinase activity -

PS Claim 4; Page 40; 48pp; English
xy

The present sequence represents a serine proteinase inhibitor containing a kn1z1t domain, designated ZKUN6. The Zkun6 polypeptides are used in the treatment and prevention of conditions associated with excessive proteinase activity. The conditions include acute pancreatitis, cardiopulmonary bypass-induced pulmonary injury, allergy-induced proinflammatory deep vein thrombosis, myocardial infarction, shock, hyperfibrinolytic haemorrhage, emphysema, rheumatoid arthritis, adult respiratory distress syndrome, chronic inflammatory bowel disease, psoriasis, inflammatory conditions, platelet function, organ preservation, and wound healing. They are also useful in the treatment of conditions arising from haemostasis imbalance, including acquired coagulopathies, primary fibrinolysis and fibrinolysis due to cirrhosis, and complications of high-dose thrombolytic therapy. Zkun6 polypeptides may also be used for inhibiting blood coagulation in mammals, and for blockade of proteolytic tissue degradation. The Zkun6 polynucleotides are used in gene therapy to treat the above diseases. Transgenic animals, engineered to express Zkun6, and knockout animals with an absence of Zkun6 function, are used to study the Zkun6 gene and the encoded protein. They are used for investigating the role of Zkun6 polypeptides in early development.

SQ Sequence 59 AA.

Query Match	12.5%;	Score 358;	DB 21;	Length 59;
Best Local Similarity	100.0%;	Pred. No. 6.5e-17;		
Matches 59;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 322 GPGDACVLPAVQGPEPCRGWEPRKMAYSPLIQOCHPEYYGGCEGNGNNFHSRESCEDCACPVP 380
|||||
Dc 1 gpqgdacvlpavqgpcprgwepcrwasyplllqqchpfyyggcegnngnnfhstrescedacvp 59

RESULT	3
AAR62523	
ID	AAR62523 standard; peptide; 560 AA

AC AAR62523;

DT 06-JUN-1995 (first entry)

Hookworm anticoagulant.

Hookworm; anticoagulant; serine protease-inhibitor; blood loss;


```

AAR92237
ID AAR92237 standard; protein; 147 AA.
AC AAR92237;
XX
XX
XX 27-SEP-1996 (first entry)
XX
XX Human wild-type urinary trypsin inhibitor.
XX
XX UTI; Kunitz domain; urinary trypsin inhibitor; elastase inhibitor;
XX recombinant protein production; yeast host cell; Pichia;
XX site-directed mutagenesis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..21
XX FT /label= signal_peptide
XX FT 26..76
XX FT Domain /label= kunitz_domain_1
XX FT Misc-difference 36..40
XX FT /label= 'P1-P4'
XX FT /note= "motif replaced by Ile-Ala-Phe-Phe-Pro
XX FT to improve elastase inhibitory activity"
XX FT Domain 78..145
XX FT /label= kunitz_domain_2
XX FT Disulfide-bond 26..76
XX FT Disulfide-bond 35..59
XX FT Disulfide-bond 51..72
XX FT Disulfide-bond 82..132
XX FT Disulfide-bond 91..115
XX FT Disulfide-bond 107..128
XX FT Modified-site 10
XX FT /label= O-linked_glycosylation
XX FT Modified-site 45
XX FT /label= N-linked_glycosylation
XX
XX W09603503-A1.
XX
XX PD 08-FEB-1996.
XX
XX PE 21-JUL-1995; 95WO-JP01449.
XX
XX PR 21-JUL-1994; 94JP-0169221.
XX
XX PA (GREC ) GREEN CROSS CORP.
XX
XX PI Goto T, Horii H, Ideno S;
XX
XX WPI; 1996-117048/12.
XX
XX PT Production of recombinant urinary trypsin inhibitor in Pichia sp. -
XX also Kunitz domain fragments of the inhibitor and new variants
XX PT having improved elastase inhibitor activity
XX
XX PS Disclosure; Fig 2; 97pp; Japanese.
XX
XX CC Recombinant urinary trypsin inhibitor (rUTI) can be produced in
XX Pichia yeast. By mutating a 5 amino acid motif (P1-P4') within Kunitz
XX domain 1 (from MGMTS to IAEFP), the resulting rUTI has improved
XX elastase inhibitory activity. The present sequence is that of
XX wild-type precursor UTI.
XX
XX SQ Sequence 147 AA;

Query Match 7.5%; Score 213; DB 17; Length 147;
Best Local Similarity 40.2%; Pred. No. 6.2e-07;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;
OY 274 CTGPTSPHLVLMHYDPQAGCGMTFARCGDGAARGFETFEACQACAGPGDACYLPAVQ 333
DB 35 cmgmstrfyf---yngtmacetlqygcmgnunfylekclqtrcrtva--acnlpdivr 88

```

```

OY 334 GPCRGWEPRWYSPILQOCHPFYGGCGEGNGNHNHRESCEDACRYP 380
DB 89 gpcratfqlwaIdavrkcvlIfpyggcggnunKfysekreycgyp 135

RESULT 6
ID AAW25928
XX AAW25928 standard; Protein; 147 AA.
XX
XX AC AAW25928;
XX
XX 11-NOV-1997 (first entry)
XX
XX Anti-trypsin inhibitor UTI.
XX
XX DE Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
XX KW neutrophil; disease; modification; site directed mutagenesis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 10
XX FT Modified-site /label= O-linked glycosylation site
XX FT Domain 22..77
XX FT Disulfide-bond 26..76
XX FT Disulfide-bond 35..59
XX FT Active-site 36
XX FT /note= "active site residue"
XX FT Disulfide-bond 51..72
XX FT Domain 78..147
XX FT /note= "Kunitz domain 2"
XX FT Disulfide-bond 82..132
XX FT Disulfide-bond 91..115
XX FT Active-site 92
XX FT /note= "active site residue"
XX FT Disulfide-bond 107..128
XX
XX JP09124700-A.
XX
XX PD 13-MAY-1997.
XX
XX PE 07-NOV-1995; 95JP-0288527.
XX
XX PR 07-NOV-1995; 95JP-0288527.
XX
XX PA (GREC ) GREEN CROSS CORP.
XX
XX WPI; 1997-316576/29.
XX
XX DR N-PSDB; AAT78950.
XX
XX PT New protease inhibitor - useful for treating diseases involving
XX PT elastase
XX
XX PS Disclosure; Fig 3; 37pp; Japanese.
XX
XX CC This is the amino acid sequence of the anti-trypsin inhibitor UTI, which
XX is a protein containing 2 Kunitz domains. The nucleotide sequence
XX encoding the protein was used to construct a novel protease inhibitor by
XX replacing the active site of the Kunitz domain 1 (see AAT79080, AAT79081
XX and AAT79083). The modified protease inhibitors are targeted to the
XX CC protease elastase, especially from neutrophils and can be used to treat
XX diseases associated with elastase. Modifications of the active site were
XX done by site directed mutagenesis.
XX
XX SQ Sequence 147 AA;

Query Match 7.5%; Score 213; DB 18; Length 147;
Best Local Similarity 40.2%; Pred. No. 6.2e-07;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

```


CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.
 XX
 XX
 SQ Sequence 352 AA;

Query Match 7.58; Score 213; DB 22; Length 352;
 Best Local Similarity 40.28; Pred. No. 1.6e-06;
 Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

OY 274 CTGPTSPHLVLMHDPQRCGCMTEPARGCDGARGFETYEACQACARGPDACVLPVQ 333
 Db 240 cmgmstsfy-----yngtsmacetlfyggcmgnnlfyetecklqctcva--acnlpivr 293

OY 334 GPCRGMPEPRMAYSPILQOCHPFYVYGCCEGNGNMFHRESCEDACPVP 380
 Db 294 gpcraflqlwafadavkqkcvlfpyggcgngnkfysekecreygcyp 340

RESULT 9
 AA014579 standard; Protein; 352 AA.
 XX

AC AA014579;

DT 24-OCT-2001 (first entry)

DE Human novel protein #450.

XX
 KW Human: novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vlnarary; nootropic;
 KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX

OS Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001: 2001WO-US02623.

XX 25-JAN-2000: 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI, 2001-451939/48.

XX N-PSDB: AAS22884.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 XX

PS Example 4; Page 862; 894pp; English.

XX The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/elicite an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.

CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.
 XX
 XX
 SQ Sequence 352 AA;

Query Match 7.58; Score 213; DB 22; Length 352;
 Best Local Similarity 40.28; Pred. No. 1.6e-06;
 Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

OY 274 CTGPTSPHLVLMHDPQRCGCMTEPARGCDGARGFETYEACQACARGPDACVLPVQ 333
 Db 240 cmgmstsfy-----yngtsmacetlfyggcmgnnlfyetecklqctcva--acnlpivr 293

OY 334 GPCRGMPEPRMAYSPILQOCHPFYVYGCCEGNGNMFHRESCEDACPVP 380
 Db 294 gpcraflqlwafadavkqkcvlfpyggcgngnkfysekecreygcyp 340

RESULT 10

AAE05095 standard; Protein; 352 AA.

XX AAE05095;

DT 18-SEP-2001 (first entry)

DE Human inter-alpha trypsin inhibitor (ITI) light chain.

XX
 KW Diagnosis; tumour; cancer; central nervous system; CNS; mab 69.31;
 KW monoclonal antibody; astrocytoma; serine protease inhibitor;
 KW primary brain tumour; brain metastasis; leukaemia; carcinoma; glioma;
 KW human; inter-alpha trypsin inhibitor light chain; ITI;
 KW plasmin inhibitor.
 XX

OS Homo sapiens.

XX Key location/Qualifiers

FT Peptide 1..19 /label= Signal_peptide

FT Protein 20..352 /label= Mature_ITI_light_chain

FT Active-site 20..32 /note= "Forms part of the active site"

FT Binding-site 215 /note= "This residue is involved in binding to ITI

FT Active-site 241..242 /note= "Forms part of the active site"

FT Active-site 297..298 /note= "Forms part of the active site"

XX WO200153835-A2.

XX 26-JUL-2001.

XX 24-JAN-2001: 2001WO-US02269.

XX 24-JAN-2000: 2000US-0491479.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Lim Y, Hixson DC;


```

Query Match 7.5%: Score 213; DB 22; Length 352;
Best Local Similarity 40.2%: Pred. No. 1.6e-06;
Matches 43; Conservative 12; Mismatches 46; Indels 6; Gaps 2;

OY 274 CTGPTSPRLVLMINDPDRGKSCFPPARGSDGAARFETYEACQACARPGDCAVLPAVQ 333
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
db 240 csmgtskyf---ygrtsmascetfuyggscmgpninlrvlekedlqrcitva--acchrlpivr 293
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 11
AAB43730
ID AAB43730 standard; Protein; 366 AA.

DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1175.

KM	Human, cancer associated gene; cancer antigen; detection; cancer;
KM	diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
KM	antidiabetic; antisthmatic; antirheumatic; antiallergic; antiviral;
KM	antifibromatous; antihypertensive; antiallergic; antibacterial; cardiant;
KM	dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KM	vasotropic; antipneumatic; antiangiogenic; gene therapy; inflammation;
KM	immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM	allergic reaction; graft versus host disease; organ rejection;
KM	haemostatic; thrombolytic; cardiovascular disorder; infection;
KM	neurological disease; drug screening.

OS Homo sapiens.

PN W0200055350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05882.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587533/55.

DR N-PSDB; AAC77939.

DR N-PSDB; AAC77939.

PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 11: Page 1797-1798; 2352pp; English.

CC AAC776030 AAC78448 encode the human cancer associated proteins given
CC in AA043398 to AA044239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; anti-neoplastic; immunomodulator;
CC antidiabetic; antiallergic; antihemorrhagic; antibacterial; antiviral;
CC antitumorigenic; antihypertensive; antidiabetic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC neurotrophic; vasotropic; antiproliferative and antitumorigenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies, and
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AA044240 represent sequences used in the exemplification of
CC the present invention.

Sequence 366 AA;

Query Match	7.5%;	Score 213;	DB 21;	Length 366;
Best Local Similarity	40.28%;	Pred. No. 1.7e-06;		
Matches 43; Conservative	12;	Mismatches 46;	Indels 6;	Gaps 2;

Qy	274	CTGPTSPHLVLMHYDPQRCGCMTPPARGCDGAARGFEITYEACQACARGPEDACVLP	AVQ	333
Db	254	cmgmmtsyf----yngtismaetfgyqgcmgngnnfvtkeclqtrtva--acnlpivr	307	307

334 GPCRGWEPRAWYSPLLQQCHPFVYGGCEGNGNMFHSRESCEDACPVP 380

Db 308 9F

RESULT 12

ID	AAW69522	standard; peptide; 147 AA
AAW69522	standard; peptide; 147 AA	

AC AAW69522;

DT 09-NOV-1998 (first entry)

DE RUTI protein SEQ ID NO:24 from WO9829453 Claim 14.

Drug; cell membrane-directed drug; phospholipid; lipid bilayer;

XX
NM CELL COLLEY, BLOOD COAGULATION, TITRATION OF VITAMIN K
NM

03 syncretic.
05 Homo sapiens.

AA W09829453-A1.
PN

09-JUL-1998

05-JAN-1998; 98WO-JP000002.

27-DEC-1996; 96JP-0359053.
(MOCH) MOCHIDA PHARM CO LTD.
Hasegawa T, Kuriyama S;
WPI: 1998-388051/33.
N-PSDB: AAV40046.
Drugs containing peptide(s) with specific affinity to phospholipid(s) - such as phosphatidyl serine, for treatment of blood coagulation, inflammatory and immunological disorders
Claim 14; Page 80-81; 117pp; Japanese.
The present invention describes drug compositions which contain as an active component a peptide which has specific affinity to particular phospholipids (such as phosphatidyl ethanolamine or phosphatidyl serine), especially to phospholipids which constitute a lipid bilayer of cellular cortex and of which the concentration in the bilayer increases in cells which are abnormal (e.g. through injury, denaturation or activation). In particular, the peptide contains a sequence having phospholipid affinity and a structure of formula (1): (A1)a-(A2)b-(A3)c, where (A1) is one of two specific sequences (see AAM69516 and AAM69517), (A2) and (A3) are TRYLRHPOSMVHQAIALR, LRYLRHPOSMVHQAIALR or MEVLGEAQNLY (see AAM69518); a = 0-5; b = 1-5; and c = 0-5. Preferred are the formulae A1-A2-A3, A2-A3, A2-A2-A3, A2-A2-A2-A3 or A2-A2 (especially A2-A2-A3, A2-A2-A2-A3 or A2-A2). The sequence is linked to a peptide such as a blood factor, especially thrombo-modulin, urina-statin or membrane cofactor protein. The drugs are used for the treatment and prevention of diseases involving blood coagulation, inflammatory and immunological disorders. The present sequence represents a peptide from the present invention.

	Query Match	7.48;	Score 212;	DB 19;	Length 147;
	Best Local Similarity	40.28;	Pred No.7.2e-07;		
Matches	43;	Conservative	11;	Mismatches	47;
				Indels	6;
				Gaps	2
OY	274 CTGPPSPHLVLMHYDRQRCCKTFPARCGDGAARGFTYEACQQAQCARGPDACVLPAAYQ	333			
	:	:	:	:	:
Db	35 cmgmstsyf---nygtsmacetlfgygcmgngnlfvtekeclqtorva--acnlpivr	88			
OY	334 GPCRGMEPRMAYSPLLOOCHPFVYGCGECENGNNHFSRESCEDACPVP	380			
	: : :				
Db	89 gpcrfflgldavdkngcnvlfpdygcgnghkfysekrecreycvp	135			
RESULT	13				
AAMW25932					
ID	AAMW25932 standard; Protein; 124 AA.				
XX					
AC	AAMW25932;				
XX					
DT	11-NOV-1997 (first entry)				
DE	New protease inhibitor Epl-d21-RPDF-52-55.				
KW	Trypsin inhibitor; Kunitz domain; protease; active site; elastase;				
RW	neutrophil; disease; modification; site directed mutagenesis.				
XX	Synthetic.				
XX					
FH	Key				
FT	Domain				
FT	/note= "Location/Qualifiers				
FT	/note= "Kunitz domain 1"				
FT	Disulfide-bond 5..55				
FT	Disulfide-bond 14..38				
FT	Misc-difference 15..19				
FT	/note= "mutated region; amino acids substituted for				

FT		wild type amino acid sequence MGMTS; this includes the active site residue of domain 1 (aa 36 of the wild type sequence AAW25928)"
FT		
FT		
Disulfide-bond	30..51	
FT	/note=	"mutated amino acid; replaces amino acid Glu in wild type sequence"
Misc-difference	31	
FT		
FT		
FT		
FT		
Misc-difference	34	
FT	/note=	"mutated amino acid; replaces amino acid Gln in wild type sequence"
Domain	57..126	
FT	/note=	"Kunitz domain 2"
Disulfide-bond	61..111	
FT		
Disulfide-bond	70..94	
FT		
Active-site	71	
FT	/note=	"active site residue"
Disulfide-bond	86..107	
XX		
JP09124700-A.		
PX		
PD		
13-MAY-1997.		
PF		
07-NOV-1995;	95JP-0288527.	
PR		
07-NOV-1995;	95JP-0288527.	
(GREC) GREEN CROSS CORP.		
DR		
N-PSTDB; AAT79080.		
New protease inhibitor - useful for treating diseases involving elastase		
Claim 5; Page 3; 37pp; Japanese.		
This is the amino acid sequence of a novel protease inhibitor based on the sequence of the anti-trypsin inhibitor UTI. The protein comprises Kunitz domains 1 and 2 of UTI, but lacks amino acids 1-21. Of the remaining amino acids, residues 1-4 (22-26 of the native sequence) are replaced by the peptide Arg-Pro-Asp-Phe. The construct also has the sequence encoding the active site residue and the adjacent amino acids (amino acids 36-40 of the wild type sequence) and amino acids 52 and 55 substituted. The modified protease inhibitors are targeted to the procase elastase, especially from neutrophils and can be used to treat diseases associated with elastase. Modifications of the active site were done by site directed mutagenesis.		
SQ		
Sequence	124 AA;	
Query Match	7.4%; Score 210; DB 18; Length 124;	
Best Local Similarity	39.0%; Pred.No. 8.le-07;	
Matches	41; Conservative 12; Mismatches 50; Indels 2; Gaps 14;	
OY	276 GPTSPHLVLNNDPORGCMTFPAFGCCGAARGFTEACQOACARGPADCVLPVAVOGSP	335
Db	12 gpciaifpyfyngscsmacqftlyvgcmgnnfnlftckedclqtcttva--acnlplivrgp	69
OY	336 CRGWEPFRNAVSYPLDQQCHPEFYVGSGCEGNNGNNFHRSRESCEDACPVP	380
Db	70 crafqlwafavdkgkcvlifpygsqdgngnkfysekreycgyvp	114
RESULT	14	
AAM25936		
ID	AAW25936 standard; Protein; 144 AA.	
AX		
AAW25936;		
DT	12-NOV-1997 (first entry)	
SUC2-Bp1-d21-RPDF-52-55 from plasmid pHH337.		

```

XX Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW neutrophil; disease; modification; site directed mutagenesis.
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..20
XX /note= "signal peptide"
XX Protein 21..144
XX /note= "mature protein"
XX Domain 21..76
XX /note= "Kunitz domain 1"
XX Disulfide-bond 25..75
XX Disulfide-bond 34..58
XX Misc-difference 35..39
XX /note= "mutated region; amino acids substituted for
XX wild type amino acid sequence MGMTS; this
XX includes the active site residue of domain 1
XX (aa 36 of the wild type sequence AAW25928)"
XX Disulfide-bond 50..71
XX Misc-difference 51..71
XX /note= "mutated amino acid; replaces amino acid Glu in
XX wild type sequence"
XX Misc-difference 54
XX /note= "mutated amino acid; replaces amino acid Glu in
XX wild type sequence"
XX Domain 77..144
XX /note= "Kunitz domain 2"
XX Disulfide-bond 81..131
XX Disulfide-bond 90..114
XX Active-site 91
XX /note= "active site residue"
XX Disulfide-bond 106..127
XX JP09124700-A.
XX
XX 13-MAY-1997.
XX
XX 07-NOV-1995; 95JP-0288527.
XX
XX 07-NOV-1995; 95JP-0288527.
XX
XX (GREC ) GREEN CROSS CORP.
XX
XX WPI: 1997-316576/29.
XX
XX N-PSDB; AAT79086.
XX
XX New protease inhibitor - useful for treating diseases involving
XX elastase
XX
XX Disclousure; Flg 30; 37pp; Japanese.
XX
XX This is the amino acid sequence of the polypeptide encoded by the insert
XX in plasmid pHH337 which comprises the novel elastase specific inhibitor
XX Epi-d21-RPDF-52-55 (AAW25932). The inhibitor sequence is linked
XX downstream of the yeast invertase (SUC2) signal peptide sequence. The
XX modified protease inhibitors are targeted to the protease elastase,
XX especially from neutrophils and can be used to treat diseases associated
XX with elastase. Modifications of the active site were done by site
XX directed mutagenesis.
XX
XX Sequence 144 AA;

```

```

Query Match 7.4%; Score 210; DB 18; Length 144;
Best Local Similarity 39.0%; Pred NO. 9.5e-07;
Matches 41; Conservative 12; Mismatches 50; Indels 2; Gaps 1;
Oy 276 GPTSPHLVLMHYDPORGCMTEPPARCGDGAARGFETYEACQACARGPDACVLPAYVGP 335
DB 32 gpciaffpyfyngtemacqftfygsgcmgnfnfvekeclqctrva--acnlpivrrp 89

```

```

Oy 336 CRGMEPRMAYSPLOQCHPFVYGCCEGNGNMFHRSCECDACPVP 380
DB 90 crafiaqlwafdkvkvclfygscgngnkfysekrecreygv 134
RESULT 15
AAW25935
ID AAW25935 standard; Protein; 145 AA.
XX
XX AAW25935;
XX
XX 11-NOV-1997 (first entry)
XX
XX Novel protease inhibitor Epi-UT1-RPDF-52-55.
XX
XX Trypsin inhibitor; Kunitz domain; protease; active site; elastase;
KW neutrophil; disease; modification; site directed mutagenesis.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 10
XX /label= O-linked glycosylation site
XX Domain 22..77
XX /note= "Kunitz domain 1"
XX Misc-difference 22..25
XX /note= "mutated amino acids: replaces wild type
XX sequence KEDS"
XX Disulfide-bond 26..76
XX Disulfide-bond 35..59
XX Misc-difference 15..19
XX /note= "mutated region; amino acids substituted for
XX wild type amino acid sequence MGMTS; this
XX includes the active site residue of domain 1"
XX Disulfide-bond 51..72
XX Misc-difference 52
XX /note= "mutated amino acid; replaces amino acid Glu in
XX wild type sequence"
XX Misc-difference 55
XX /note= "mutated amino acid; replaces amino acid Glu in
XX wild type sequence"
XX Domain 78..147
XX /note= "Kunitz domain 2"
XX Disulfide-bond 82..132
XX Disulfide-bond 91..115
XX Active-site 92
XX /note= "active site residue"
XX Disulfide-bond 107..128
XX JP09124700-A.
XX
XX 13-MAY-1997.
XX
XX 07-NOV-1995; 95JP-0288527.
XX
XX 07-NOV-1995; 95JP-0288527.
XX
XX (GREC ) GREEN CROSS CORP.
XX
XX WPI: 1997-316576/29.
XX
XX N-PSDB; AAT79083.
XX
XX New protease inhibitor - useful for treating diseases involving
XX elastase
XX
XX Claim 8; Page 4; 37pp; Japanese.
XX
XX This is the amino acid sequence of a protease inhibitor construct based
XX on the anti-trypsin inhibitor UT1 (AAW25928). The novel inhibitor
XX contains the Kunitz domains 1 and 2 of UT1 with several amino acid
XX substitutions: amino acids 22-25 of the wild type protein (KEDS) are
XX replaced by the sequence RPDF; the active site sequence and adjacent
XX amino acids (amino acids 36-40 of the wild type sequence) are replaced by

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:53:19 ; Search time 78.18 Seconds
(without alignments)
11.226 Million cell updates/sec

Title: US-09-819-136-2_COPY_50_55

Perfect score: 6
Sequence: 1 ERECSR 6

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_17:*

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	83.3	35	5 Q9U780	Q9U780 boophilus a
2	5	83.3	121	12 Q9I239	Q9I239 bacillus su
3	5	83.3	128	2 Q3I739	Q3I739 bacillus su
4	5	83.3	153	9 Q9MCM0	Q9MCM0 streptococ
5	5	83.3	184	5 Q9N411	Q9N411 caenorhabdi
6	5	83.3	193	1 Q9HMG4	Q9HMG4 halobacteri
7	5	83.3	196	4 Q14557	Q14557 homo sapien
8	5	83.3	205	4 Q9NXX68	Q9NXX68 homo sapien
9	5	83.3	213	10 Q9LEP8	Q9LEP8 brassica na
10	5	83.3	216	12 Q96670	Q96670 chicken ane
11	5	83.3	216	12 Q41447	Q41447 chicken ane
12	5	83.3	216	12 Q9IZU7	Q9IZU7 chicken ane
13	5	83.3	216	12 Q9IEZ5	Q9IEZ5 chicken ane
14	5	83.3	216	12 Q9DXA8	Q9DXA8 chicken ane
15	5	83.3	216	12 Q9DWA4	Q9DWA4 chicken ane
16	5	83.3	216	12 Q9DWA0	Q9DWA0 chicken ane
17	5	83.3	216	12 Q9DWM7	Q9DWM7 chicken ane
18	5	83.3	216	12 Q9DGM8	Q9DGM8 chicken ane
19	5	83.3	267	5 Q21692	Q21692 caenorhabdi

20	5	83.3	270	2 Q9WZQ2	Q9WZQ2 thermotoga
21	5	83.3	286	1 Q9YAJ9	Q9YAJ9 aeropyrum p
22	5	83.3	290	2 Q52611	Q52611 proteus vul
23	5	83.3	290	2 Q9I331	Q9I331 serralia ma
24	5	83.3	299	10 Q9LGL6	Q9LGL6 oryza sativ
25	5	83.3	329	3 Q9P708	Q9P708 schizosacch
26	5	83.3	335	11 Q9JUG8	Q9JUG8 mus musculi
27	5	83.3	337	2 Q9JRF0	Q9JRF0 neisseria m
28	5	83.3	354	11 Q9DAC9	Q9DAC9 mus musculi
29	5	83.3	370	1 Q58094	Q58094 pyrococcus
30	5	83.3	376	11 Q61055	Q61055 mus musculi
31	5	83.3	395	2 Q9ADK5	Q9ADK5 streptomyce
32	5	83.3	444	2 Q67580	Q67580 aquifex aeo
33	5	83.3	458	2 Q92730	Q92730 chlamydia p
34	5	83.3	458	2 Q9I624	Q9I624 pseudomonas
35	5	83.3	556	4 Q9NVE0	Q9NVE0 homo sapien
36	5	83.3	565	10 Q9M195	Q9M195 arabidopsis
37	5	83.3	568	5 Q96182	Q96182 plasmodium
38	5	83.3	585	10 Q9AUD0	Q9AUD0 sesamum ind
39	5	83.3	597	10 Q9EJY3	Q9EJY3 arabidopsis
40	5	83.3	646	2 Q9L076	Q9L076 streptomyce
41	5	83.3	665	2 Q9ZC03	Q9ZC03 rickettsia
42	5	83.3	665	2 Q60966	Q60966 leishmania
43	5	83.3	667	2 P71538	P71538 mycobacteri
44	5	83.3	669	2 Q9A6B5	Q9A6B5 caulobacter
45	5	83.3	669	6 Q28085	Q28085 bos taurus

ALIGNMENTS

RESULT 1					
Q9U780		PRELIMINARY:	PRT:	35 AA.	
AC Q9U780;					
DT 01-MAY-2000 (TREMBlrel. 13, Created)					
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)					
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)					
DE BA99 (FRAGMENT).					
GN BA99.					
OS Boophilus annulatus.					
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;					
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.					
OX NCBI_Taxid=34611;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Garcia-Garcia J.C., Gonzalez I.L., Gonzalez D.M., Valdes M.,					
RA Mendez L., Lambert J., D'Agostino B., Clitroni D., Fragoso H.,					
RA Ortiz M., Rodriguez M., de la Fuente J.,					
RT "Sequence variations in the Boophilus microplus Bm86 locus and					
RT implications for immunoprotection in cattle vaccinated with this					
RT antigen."					
RT Exp. Appl. Acarol. 0:0-0(1999).					
DR EMBL: AF150897; AAF01360.1; -.					
FT NON_TER	1				
FT NON_TER	35				
SO SEQUENCE	35 AA;	4155 MW;	C8E70E7483F70961 CRC64;		
Query Match		83.3%;	Score 5;	DB 5;	
Best Local Similarity		100.0%;	Pred. No. 10;		
Matches	5;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY 2 RECSR 6					
Db 27 RECSR 31					
RESULT 2					
Q9I239		PRELIMINARY:	PRT:	121 AA.	
ID Q9I239;					
AC Q9I239;					
DT 01-OCT-2000 (TREMBlrel. 15, Created)					

```

DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE VP3 (FRAGMENT).
OS chicken anemia virus.
OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
OX NCBI_TaxId=12618;
RN
  [1]
  SEQUENCE FROM N.A.
RA Yang S., Lee J., Chiu C., Chueh L.;
RT "The cloning and sequencing of VP3 of Taiwan 11an strain of chicken
  anemia virus.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF242190; AAF44705.1; -.
FT NON_TER 1
FT NON_TER 121
SQ SEQUENCE 121 AA; 14339 MW; 2E3DE2FC0D3804D6C CRC64;

Query Match
Best Local Similarity 100.0%; Score 5; DB 12; Length 121;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 49 RECSR 53

RESULT 3
O31739 PRELIMINARY; PRT; 128 AA.
AC O31739;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE Y10D PROTEIN.
GN Y10D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilli/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1423;
  [1]
  SEQUENCE FROM N.A.
RP STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bojoltin A., Borcher S.,
RA Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuno V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Filtz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Meliade R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Portoullik S., Prescott A.M.,
RA Pressecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocho E., Roche B., Rose M., Sadle Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weltenegeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus

```

```

RT subtilis.";
RL Nature 390:249-256(1997).
RN
  [2]
  SEQUENCE FROM N.A.
RP STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RT Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z69112; CAB13474.1; -.
KW Complete proteome.
SQ SEQUENCE 128 AA; 15039 MW; 7D6B49BD1CFAE706 CRC64;

Query Match
Best Local Similarity 100.0%; Score 5; DB 2; Length 128;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
DB 33 ERECS 37

RESULT 4
O9MCM0 PRELIMINARY; PRT; 153 AA.
AC O9MCM0;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE ORE10.
GN ORE10.
OS Streptococcus thermophilus bacteriophage 7201.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxId=112023;
  [1]
  SEQUENCE FROM N.A.
RP MEDLINE=20088830; PubMed=10620678;
RA Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;
RA van Sinderen D.;
RT "Identification of four loci isolated from two Streptococcus
  thermophilus phage genomes responsible for mediating bacteriophage
  resistance.";
RL FEMS Microbiol. Lett. 182:271-277(2000).
RN
  [2]
  SEQUENCE FROM N.A.
RP Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF145054; AAF43503.1; -.
SQ SEQUENCE 153 AA; 18008 MW; 817027FE7E42816E CRC64;

Query Match
Best Local Similarity 100.0%; Score 5; DB 9; Length 153;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 33 RECSR 37

RESULT 5
O9N411 PRELIMINARY; PRT; 184 AA.
AC O9N411;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Y11F9AL.14 PROTEIN.
GN Y11F9AL.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
  [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode *C. elegans*: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 283:2012-2018(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC024200; AAF36009.2;
DR InterPro: IPR001451; Hexapep_transf.
DR Pfam: PF00132; hexapep; 5.
SQ SEQUENCE 184 AA; 20307 MW; C388C1E58113478F CRC64;

Query Match 83.3%; Score 5; DB 5; Length 184;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 RECSR 6
Db 123 RECSR 127

RESULT 6
Q9HMG4 PRELIMINARY; PRT; 193 AA.
AC Q9HMG4;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE VNG2554H.
GN VNG2554H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20504483; PubMed-11016950;
RA Ng W.V., Kennedy S.P., Mahalax G.G., Bergquist B., Pan M.,
Shukla H.D., Lasky S.R., Ballig N.S., Thorsson V., Sirogna J.,
Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
Edhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005131; AAG20607.1;
DR InterPro: IPR002818; ThLJ.
DR Pfam: PF01965; ThLJ; 1.
KW Complete proteome.
SQ SEQUENCE 193 AA; 20245 MW; 16291CEE76BBFEED CRC64;

Query Match 83.3%; Score 5; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ERECS 5
Db 172 ERECS 176

RESULT 7
ID 014557 PRELIMINARY; PRT; 196 AA.
AC 014557;

DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE F25965_1 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhart-Schultz K.,
RA Garcia E., Kyle A., Ramirez M., Stillwegen S., Gaines J., Dangnan L.,
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA Carraro A.V.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002398; AAB81199.1;
FT NON_TER 1
SQ SEQUENCE 196 AA; 22942 MW; B828E433289933E1E CRC64;

Query Match 83.3%; Score 5; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ERECS 5
Db 81 ERECS 85

RESULT 8
Q9NX68 PRELIMINARY; PRT; 205 AA.
AC Q9NX68;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CDNA FLJ20413 FIS, CLONE KAT02170.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Oobayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono T., Sugano S.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK00420; BAA91152.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 205 AA; 22275 MW; 64BFBAB3996AF670 CRC64;

Query Match 83.3%; Score 5; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 RECSR 6
Db 53 RECSR 57

RESULT 9
Q9LEP8 PRELIMINARY; PRT; 213 AA.
AC Q9LEP8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 23.8 KDA PROTEIN (FRAGMENT).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, N-O-9;
 RA Bowers N.L., Trick M.;
 RT "Microsytely at the FCA region between Arabidopsis thaliana and
 Brassica napus."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ293574; CAB98167.1;
 KW Hypothetical protein.
 FT NON_TER
 FT 213
 SQ SEQUENCE 213 AA; 23759 MW; F47AC8192DF8BD72 CRC64;

Query Match 83.3%; Score 5; DB 10; Length 213;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 Db 70 RECSR 74

RESULT 10
 O96670 PRELIMINARY; PRT; 216 AA.
 AC O96670;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
 DE COMPLETE GENOME: VP2, VP3 AND VP1 GENES, COMPLETE CDS.
 OS Chicken anaemia virus.
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=12618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-704;
 RA Hamouleh G., Cross G., Dixon R.J.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U65414; AAB06180.1;
 SQ SEQUENCE 216 AA; 24124 MW; 9346D9848CE8CA03 CRC64;

Query Match 83.3%; Score 5; DB 12; Length 216;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 Db 85 RECSR 89

RESULT 11
 O41447 PRELIMINARY; PRT; 216 AA.
 ID O41447
 AC O41447;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE CLONED ISOLATE 10, COMPLETE GENOME.
 OS chicken anaemia virus.
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=12618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CLONED ISOLATE 10;
 RX MEDLINE=92296898; Pubmed=1605740;
 RA Meehan B.M., Todd D., Creelan J.L., Earle J.A.P., Hoey E.M.,
 RA McNulty M.S.;
 RT "Characterization of viral DNAs from cells infected with chicken
 anaemia agent: sequence analysis of the cloned replicative form and

RT transfection capabilities of cloned genome fragments.";
 RL Arch. Virol. 124:301-319(1992).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CLONED ISOLATE 10;
 RA Todd D., Connor T.J., Calvert V., Creelan J.L., Meehan B.M.,
 RA McNulty M.S.;
 RL Avian Pathol. 24:171-187(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CLONED ISOLATE 10;
 RX MEDLINE=96001356; Pubmed=9343191;
 RA Meehan B.M., Todd D., Creelan J.L., Connor T.J., McNulty M.S.;
 RT "Investigation of the attenuation exhibited by a molecularly cloned
 chicken anaemia virus isolate by utilizing a chimeric virus approach.";
 RL J. Virol. 71:8362-8367(1997).

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CLONED ISOLATE 10;
 RA Meehan B.M.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U66304; AAC58476.1;
 SQ SEQUENCE 216 AA; 24053 MW; BB7C39E09D6F1EBE CRC64;

Query Match 83.3%; Score 5; DB 12; Length 216;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 Db 85 RECSR 89

RESULT 12
 O91ZU7 PRELIMINARY; PRT; 216 AA.
 ID O91ZU7
 AC O91ZU7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE VP2.
 OS chicken anaemia virus.
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=12618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CAU269/7;
 RA Brown H.K., Browning G.F., Scott P.C., Grabb B.S.;
 RT "A full-length infectious clone of a pathogenic Australian isolate of
 chicken anaemia virus."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF227982; AAF34787.1;
 SQ SEQUENCE 216 AA; 24111 MW; F83EB3D07B97F752 CRC64;

Query Match 83.3%; Score 5; DB 12; Length 216;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 Db 85 RECSR 89

RESULT 13
 O91EZ5 PRELIMINARY; PRT; 216 AA.
 ID O91EZ5
 AC O91EZ5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VP2 PROTEIN.

GN VP2.
 OS Chicken anemia virus (German isolate Cuxhaven-1) (CAV), and
 OS chicken anemia virus.
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=73475, 12618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=chicken anemia virus; STRAIN=CUXHAVEN-1;
 RA Scott A.N.J., Todd D.;
 RT "Characterisation of a Chicken Anaemia Virus variant population that
 resists neutralisation with a group-specific monoclonal antibody.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=chicken anemia virus; STRAIN=CUX-1;
 RA Scott A.N.J., McNulty M.S., Todd D.;
 RT "Characterisation of a chicken anaemia virus variant population that
 resists neutralisation with a group-specific monoclonal antibody.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ297684; CAC14763.1; -;
 DR EMBL: AJ297678; CAC14745.1; -;
 DR EMBL: AJ297679; CAC14748.1; -;
 DR EMBL: AJ297680; CAC14751.1; -;
 DR EMBL: AJ297683; CAC14760.1; -;
 SQ SEQUENCE 216 AA; 24050 MW; 2670206C1D9F9210 CRC64;

Query Match 83.3%; Score 5; DB 12; Length 216;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 |||||
 DB 85 RECSR 89

RESULT 14
 O9DXA8 PRELIMINARY; PRT; 216 AA.
 AC O9DXA8;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE VP2.
 OS chicken anemia virus.
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=12618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DEL-ROS;
 RA Spackman E., Rosenberger J.K.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF313470; AAG34179.1; -;
 SQ SEQUENCE 216 AA; 24196 MW; A9D789DEA7210C0 CRC64;

Query Match 83.3%; Score 5; DB 12; Length 216;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 |||||
 DB 85 RECSR 89

RESULT 15
 O9DXA4 PRELIMINARY; PRT; 216 AA.
 AC O9DXA4;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE VP2 PROTEIN.
 GN VP2.
 OS chicken anemia virus.
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=12618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CUX-1;
 RA Scott A.N.J., McNulty M.S., Todd D.;
 RT "Characterisation of a chicken anaemia virus variant population that
 resists neutralisation with a group-specific monoclonal antibody.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ297681; CAC14754.1; -;
 SQ SEQUENCE 216 AA; 24038 MW; D25B76E098BDDC65 CRC64;

Query Match 83.3%; Score 5; DB 12; Length 216;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 |||||
 DB 85 RECSR 89

Search completed: February 26, 2002, 01:53:21
 Job time: 445 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:53:55 ; Search time 24.25 Seconds

(without alignments)
9.072 Million cell updates/sec

Title: US-09-819-136-2_COPY_50_55

Perfect score: 6
Sequence: 1 ERECSR 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5	83.3	216	VP2_CAV26	P54092 chicken ane
2	5	83.3	216	VP2_CAV82	P54093 chicken ane
3	5	83.3	216	VP2_CAVC1	099151 chicken ane
4	5	83.3	329	YFV6_SCHPO	09P768 schizosach
5	5	83.3	337	YKPL_YEAST	P36059 saccharomyc
6	5	83.3	456	PRTC_BOVIN	P00745 bos taurus
7	5	83.3	506	PYCA_ARCFU	030019 archaeoglob
8	5	83.3	514	CGW1_DROME	09V911 drosophila
9	5	83.3	571	PTL1_CHLTR	084340 chlamydia t
10	5	83.3	631	ADAS_DROME	09V778 drosophila
11	5	83.3	654	BCCA_MYCTU	P46401 mycobacteri
12	5	83.3	703	PCCA_HUMAN	P05165 homo sapien
13	5	83.3	704	PCCA_RAT	P14882 rattus norv
14	5	83.3	1093	P14K_DICDI	P54677 dictyostel
15	5	83.3	1234	CEAH_MOUSE	P06909 mus musculu
16	5	83.3	2353	CCAH_HUMAN	095180 homo sapien
17	4	66.7	67	YMS4_ARCFU	028049 archaeoglob
18	4	66.7	76	RSJ7_YEAST	P05759 saccharomyc
19	4	66.7	88	MEG1_MOUSE	061845 mus musculu
20	4	66.7	98	NI8M_BOVIN	002370 bos taurus
21	4	66.7	98	NI8M_HUMAN	043678 homo sapien
22	4	66.7	99	YCIK_ECOLI	P58094 escherichia
23	4	66.7	105	YGRW_MYCEC	P24650 micromonosp
24	4	66.7	111	RS17_ARCFU	028363 archaeoglob
25	4	66.7	132	YDHL_ECOLI	P76188 escherichia
26	4	66.7	135	WAP_PIG	O46655 sus scrofa
27	4	66.7	134	WAP_MOUSE	P01113 mus musculu
28	4	66.7	135	HNS_BUCAI	P57360 buchnera ap
29	4	66.7	156	YH07_THEMA	O9X223 thermotoga
30	4	66.7	157	NUGC_STNY3	P19155 synechocyst
31	4	66.7	157	UBI4_HUMAN	P11441 homo sapien
32	4	66.7	157	UBI4_MOUSE	P21126 mus musculu
33	4	66.7	161	Y255_BUCAI	P57343 buchnera ap

34	4	66.7	172	1	COTX_BACSU	O08313 bacillus su
35	4	66.7	175	1	VP6_BMYVE	P09511 beet wester
36	4	66.7	180	1	TRIC_TRIVU	O29147 trichosurus
37	4	66.7	184	1	YDJA_HAETN	P45244 haemophilus
38	4	66.7	195	1	NUGM_ACAKA	O37383 acanthamoeb
39	4	66.7	200	1	E1A_ADEMI	P12534 mouse adeno
40	4	66.7	200	1	RUVA_CHLMU	O9P350 chlamydia m
41	4	66.7	200	1	RUVA_CHLTR	O84509 chlamydia t
42	4	66.7	207	1	RUVA_CHLPN	O92714 chlamydia p
43	4	66.7	213	1	YFEL_ECOLI	P76543 escherichia
44	4	66.7	215	1	FIMI_ECOLI	P39264 escherichia
45	4	66.7	226	1	AROD_LACIA	O9C139 lactococcus

ALIGNMENTS

RESULT 1	VP2_CAV26	STANDARD;	PRT;	216 AA.
ID	VP2_CAV26			
AC	P54092;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	PROTEIN VP2.			
OS	Chicken anemia virus (USA isolate 26p4) (CAV).			
OC	viruses; ssDNA viruses; Circoviridae; Circovirus.			
OX	NCBI_Taxid=73477;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE:91341490; PubMed:1908516;			
RA	Claessens J.A.J., Schrier C.C., Mockett A.P.A., Jagt E.H.J.M.,			
RA	Sondermeijer P.J.A.;			
RT	"Molecular cloning and sequence analysis of the genome of chicken			
RT	anaemia agent.";			
RL	J. Gen. Virol. 72:2003-2006(1991).			
CC	-I- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF			
CC	INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: D10068; BAA00959.1; -			
DR	SEQUENCE 216 AA; 24066 MW; CD4869EF88BDDC72 CRC64;			
QY	2 RECSR 6			
QY				
DB	85 RECSR 89			
Query Match	83.3%; Score 5; DB 1; Length 216;			
Best local similarity	100.0%; Pred. No. 9.5;			
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
RESULT 2	VP2_CAV82	STANDARD;	PRT;	216 AA.
ID	VP2_CAV82			
AC	P54093;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	PROTEIN VP2.			
OS	Chicken anemia virus (Japanese isolate 82-2) (CAV).			
OC	viruses; ssDNA viruses; Circoviridae; Circovirus.			
OX	NCBI_Taxid=73476;			
RP	[1]			
RP	SEQUENCE FROM N.A.			

RX MEDLINE-95297149; PubMed-7778281;
 RA Kato A., Fujino M., Nakamura T., Ishihama A., Otaki Y.;
 RT "Gene organization of chicken anemia virus.";
 RL Virology 209:480-488(1995).
 CC -1- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
 CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D31965; BAA06732.1; -
 DR EMBL: D31965; BAA06732.1; -
 CC SEQUENCE 216 AA; 24138 MW; E23169F592BDE52 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 216;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 DB 85 RECSR 89

RESULT 3
 VP2_CAVC1 STANDARD; PRT; 216 AA.

AC 099151; P54091;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PROTEIN VP2.
 OS Chicken anemia virus (German isolate Cuxhaven-1) (CAV), and
 OS Chicken anemia virus (USA isolate CIA-1) (CAV).
 OC Viruses; ssDNA viruses; Circoviridae; Circovirus.
 OX NCBI_TaxID=73475, 73478;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-CUXHAVEN-1;
 RX MEDLINE-91237831; PubMed-1851873;
 RA Noteborn M.H.M., de Boer G.F., van Roozelaar D.J., Karremans C.,
 RA Kienburg O., Vos J.G., Jeurissen S.H.M., Hoebe R.C., Zanema A.,
 RA Koch G., van Ormondt H., van der Eb A.J.;
 RT "Characterization of cloned chicken anemia virus DNA that contains
 RT all elements for the infectious replication cycle.";
 RL J. Virol. 65:3131-3139(1991).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-CUXHAVEN-1;
 RX MEDLINE-92296898; PubMed-1605740;
 RA Meenan B.M., Todd D., Creelan J.L., Earle J.A.P., Hoey E.M.,
 RA McNulty M.S.;
 RT "Characterization of viral DNAs from cells infected with chicken
 RT anemia agent: sequence analysis of the cloned replicative form and
 RT transfection capabilities of cloned genome fragments.";
 RL Arch. Virol. 124:301-319(1992).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-CIA-1;
 RX MEDLINE-97126092; PubMed-8971016;
 RA Renshaw R.W., Solne C., Weinkle T., O'Connell P.H., Ohashi K.,
 RA Watson S., Lucio B., Harrington S., Schat K.A.;
 RT "A hypervariable region in VP1 of chicken infectious anemia virus
 RT mediates rate of spread and cell tropism in tissue culture.";
 RL J. Virol. 70:8872-8878(1996).
 RN [4]

RP REVISION TO 24.
 RA Renshaw R.W.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
 CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M55918; AAA91822.1; -
 DR EMBL: M55918; AAA91822.1; -
 DR EMBL: M81223; AAA42882.1; -
 CC EMBL: L14767; AAD09422.1; -
 FT CONFLICT 153 153 V -> A (IN REF. 1).
 FT CONFLICT 187 187 D -> N (IN REF. 2).
 CC SEQUENCE 216 AA; 24138 MW; F82B69EF8BDE52 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 216;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 DB 85 RECSR 89

RESULT 4
 YFV6_SCHPO STANDARD; PRT; 329 AA.

AC 09P708;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 38.5 KDA PROTEIN C1834.06C IN CHROMOSOME I.
 GN SPAC1834.06C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONG TO THE MO25 FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: AL157734; CAB5774.1; -
 DR EMBL: AL157734; CAB5774.1; -
 KW Hypothetical protein.
 CC SEQUENCE 329 AA; 38521 MW; 073DD0607A64C952 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 329;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
 DB 144 RECSR 148

RESULT 5
 YKPL_YEAST

```

ID YKPI_YEAST STANDARD; PRT; 337 AA.
AC P36059;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 37.4 KDA PROTEIN IN GPWI-MCRI INTERGENIC REGION.
GN YKL151C OR YKL606.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378720; PubMed=8091859;
RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
RT "DNA sequencing of a 36.2 kb fragment located between the FAST and
RT LAP loci of chromosome XI of Saccharomyces cerevisiae.";
RL Yeast 10:S35-S40(1994).
RN [2]
RP SIMILARITY.
RA Bork P.;
RL Unpublished observations (Aug-1994).
CC -1- SIMILARITY: BELONGS TO THE UPF0031 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z26877; CAA81502.1; -
DR EMBL: Z28151; CAA81992.1; -
DR PIR: S37799; S37799.
DR PIR: S44576; S44576.
DR SGD: S0001634; YKL151C.
DR InterPro: IPR000631; UPF0031.
DR Pfam: PF01256; UPF0031.1.
DR PROSITE: PS01049; UPF0031_1; 1.
DR PROSITE: PS01050; UPF0031_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 337 AA; 37357 MW; 472C46DB6956244E CRC64;

Query Match 83.3%; Score 5; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
DB 305 RECSR 309

RESULT 6
PRTC BOVIN STANDARD; PRT; 456 AA.
AC P00745;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE VITAMIN-K DEPENDENT PROTEIN C PRECURSOR (EC 3.4.21.69)
DE (AUTOPROTHROMBIN IIA) (ANTICOAGULANT PROTEIN C) (BLOOD COAGULATION
DE FACTOR XIV) (FRAGMENT).
GN PROC.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=85014826; PubMed=6091100;
RA Long G.L., Balagaja R.M., McGillivray R.T.A.;
RT "Cloning and sequencing of liver cDNA coding for bovine protein C.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5653-5656(1984).
RN [2]
RP SEQUENCE OF 40-194.
RX MEDLINE=83007325; PubMed=6896876;
RA Fernlund P., Stenflo J.;
RT "Amino acid sequence of the light chain of bovine protein C.";
RL J. Biol. Chem. 257:12170-12179(1982).
RN [3]
RP REVISION TO 110.
RX MEDLINE=83169769; PubMed=6572939;
RA Drakenberg T., Fernlund P., Roepstorff P., Stenflo J.;
RT "Beta-hydroxyaspartic acid in vitamin K-dependent protein C.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:1802-1806(1983).
RN [4]
RP SEQUENCE OF 197-456.
RX MEDLINE=83007326; PubMed=6896877;
RA Stenflo J., Fernlund P.;
RT "Amino acid sequence of the heavy chain of bovine protein C.";
RL J. Biol. Chem. 257:12180-12190(1982).
RN [5]
RP PROCESSING, AND CALCIUM-BINDING DATA.
RX MEDLINE=83213513; PubMed=6304092;
RA Esmon N.L., Debault L.E., Esmon C.T.;
RT "Proteolytic formation and properties of gamma-carboxyglutamic acid-
RL domainless protein C.";
RN J. Biol. Chem. 258:5548-5553(1983).
RN [6]
RP PROCESSING, AND CALCIUM-BINDING DATA.
RX MEDLINE=83213514; PubMed=6406503;
RA Johnson A.E., Esmon N.L., Lane T.M., Esmon C.T.;
RT "Structural changes required for activation of protein C are induced
RT by Ca2+ binding to a high affinity site that does not contain gamma-
RT carboxyglutamic acid.";
RL J. Biol. Chem. 258:5554-5560(1983).
CC -1- FUNCTION: PROTEIN C IS A VITAMIN K-DEPENDENT SERINE PROTEASE THAT
CC REGULATES BLOOD COAGULATION BY INACTIVATING FACTORS VA AND VIIIA
CC IN THE PRESENCE OF CALCIUM IONS AND PHOSPHOLIPIDS.
CC -1- CATALYTIC ACTIVITY: DEGRADATION OF BLOOD COAGULATION FACTORS VA
CC AND VIIIA.
CC -1- SUBUNIT: SYNTHESIZED AS A SINGLE CHAIN PRECURSOR, WHICH IS CLEAVED
CC INTO A LIGHT CHAIN AND A HEAVY CHAIN HELD TOGETHER BY A DISULFIDE
CC BOND. THE ENZYME IS THEN ACTIVATED BY THROMBIN, WHICH CLEAVES A
CC TETRADECAPEPTIDE FROM THE AMINO END OF THE HEAVY CHAIN. THIS
CC REACTION, WHICH OCCURS AT THE SURFACE OF ENDOTHELIAL CELLS, IS
CC STRONGLY PROMOTED BY THROMBOMODULIN.
CC -1- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -1- PTM: THE VITAMIN K-DEPENDENT, ENZYMAIC CARBOXYLATION OF SOME
CC GLU RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND CALCIUM.
CC -1- MISCELLANEOUS: CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO
CC ANOTHER SITE, BEYOND THE GLA DOMAIN. THIS GLA-INDEPENDENT BINDING
CC SITE IS NECESSARY FOR THE RECOGNITION OF THE
CC THROMBIN-THROMBOMODULIN COMPLEX.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K02435; AAA30685.1; -
DR PIR: A00928; KXBO.
DR HSSP: P04070; IAUT.
DR MEROPS: S01.218; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.

```

DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001284; Trypsin.
 DR InterPro: IPR000294; VitK_dep_GLA.
 DR Pfam: PF00008; EGF_1.
 DR Pfam: PF00594; gla; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SM00181; EGF_2.
 DR SMART: SM00069; GLA; 1.
 DR SMART: SM00020; TRYP_SPE; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
 DR PROSITE: PS00240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Blood coagulation; Glycoprotein; Plasma; Serine protease; Liver;
 KW Gamma-carboxyglutamic acid; Calcium binding; Vitamin K; Hydroxylation;
 KW EGF-like domain; Repeat; Endothelial cell; Hydrolyase; Signal.
 FT NON_TER 1
 FT SIGNAL 1
 FT PROPEP 30
 FT CHAIN 40
 FT CHAIN 194
 FT CHAIN 197
 FT PEPTIDE 197
 FT DOMAIN 94
 FT DOMAIN 133
 FT DOMAIN 173
 FT DOMAIN 211
 FT MOD_RES 45
 FT MOD_RES 46
 FT MOD_RES 53
 FT MOD_RES 55
 FT MOD_RES 58
 FT MOD_RES 59
 FT MOD_RES 62
 FT MOD_RES 64
 FT MOD_RES 65
 FT MOD_RES 68
 FT MOD_RES 74
 FT MOD_RES 110
 FT ACT_SITE 252
 FT ACT_SITE 298
 FT ACT_SITE 397
 FT ACT_SITE 397
 FT DISULFID 56
 FT DISULFID 89
 FT DISULFID 98
 FT DISULFID 102
 FT DISULFID 117
 FT DISULFID 119
 FT DISULFID 137
 FT DISULFID 148
 FT DISULFID 144
 FT DISULFID 159
 FT DISULFID 172
 FT DISULFID 180
 FT DISULFID 237
 FT DISULFID 382
 FT DISULFID 393
 FT CARBOHYD 136
 FT CARBOHYD 289
 FT CARBOHYD 350
 FT CARBOHYD 366
 FT CARBOHYD 366
 FT VARIANT 82
 FT VARIANT 82
 FT CONFLICT 455
 FT CONFLICT 456
 SQ SEQUENCE 456 AA; 51407 MW; 51407 MW; CAAPF833F894C209 CRC64;

VP -> PV (IN REF. 4).
 CAAPF833F894C209 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 456;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
 DB 53 ERECS 57

RESULT 7
 PYCA_ARCFU STANDARD; PRT; 506 AA.
 AC 030019;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PYRUVATE CARBOXYLASE SUBUNIT A (EC 6.4.1.1) (PYRUVIC CARBOXYLASE A).
 GN PYCA OR AF0220.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OX NCBI_TaxID=2234;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., Hickey E.K., Peterson J.D.,
 RA Richardson K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kervilavage A.R., Graham D.E., Kyprides N.C.,
 RA Fletschmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöck A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION
 INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
 ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE CARBOXYL
 GROUP TO PYRUVATE IN THE SECOND.
 CC -1- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3)(-) = ADP +
 ORTHOPHOSPHATE + OXALOACETATE.
 CC -1- COFACTOR: ATP, MAGNESIUM (OR MANGANESE OR COBALT), PYRUVATE AND
 BICARBONATE (BY SIMILARITY).
 CC -1- PATHWAY: GLUCONEOGENESIS.
 CC -1- SUBUNIT: HETEROOCTAMER OF FOUR A AND FOUR B SUBUNITS (BY
 SIMILARITY).
 CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPAMIDE TRANSFERASES
 AND CARBAMYL PHOSPHATE SYNTHETASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE001090; AAF91012.1; -.
 DR HSSP: P24182; IBNC.
 DR TIGR: AF0220; -.
 DR InterPro: IPR000901; CPSSase.
 DR Pfam: PF00289; CPSSase_L_chain; 1.
 DR PROSITE: PS00866; CPSSase_1; 1.
 DR PROSITE: PS00867; CPSSase_2; FALSE_NEG.
 KW Ligase; Multifunctional enzyme; Gluconeogenesis; Magnesium; Pyruvate;
 KW ATP-binding; Complete proteome.
 FT NP_BIND 161 166 ATP (POTENTIAL).
 FT ACT_SITE 290 290 POTENTIAL.
 SQ SEQUENCE 506 AA; 57438 MW; 39B765F319235AD1 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 506;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5

Db 225 ERECS 229

```

RESULT 8
C6M1_DROME STANDARD: PRT; 514 AA.
ID C6M1_DROME
AC 09V9L1;
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE CYTOCHROME P450 6W1 (EC 1.14.-.-) (CYP6W1).
GN CYP6W1 OR CG8345.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chame M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abdayeni A., An H.-J., Andrews-Planckoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borjova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Jostali D., Houston K.A., Howland T.J., Wei M.-H., Ibergan C.,
RA Jostali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Metulov G., Milshina N.V., Modarry C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rehbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitslks R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE003784; AAP57277.1; -
DR Flybase: FBgn003065; Cyp6w1.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Hypothetical protein.
FT BINDING 450 450 HEME (BY SIMILARITY).
SQ SEQUENCE 514 AA; 59320 MW; 1CA0E9F2692825E5 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 514;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ERECS 5
Db 374 ERECS 378

RESULT 9
PT1_CHLTR STANDARD: PRT; 571 AA.
ID PT1_CHLTR
AC 084340;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHONOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (EC 2.7.3.9)
DE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I).
OS Chlamydia trachomatis.
GN Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/W-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHONOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
CC PHOSPHONOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC ENZYME I IS COMMON TO ALL PTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: PHOSPHONOLPYRUVATE + PROTEIN HISTIDINE -
CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001306; AAC67931.1; -
DR InterPro: IPR000121; PEP_utilizers.
DR Pfam: PF00391; PEP_utilizers; 2.
DR ProDom: PD000940; PEP_utilizers; 1.
DR PROSITE: PS00740; PEP_ENZYMES_PHOS_SITE; FALSE_NEG.
DR PROSITE: PS00742; PEP_ENZYMES_2; FALSE_NEG.
KW Phosphotransferase system; Transferase; Kinase; Sugar transport;
KW Phosphorylation; Complete proteome.

```

```

FT MOD.RES 207 207 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 571 AA: 63786 MW: 73EBFEE7C457067 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 571;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECS 6
DB 547 RECS 551

RESULT 10
ADAS_DROME STANDARD: PRT: 631 AA.
ID 09V778:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALKYLIDHYDROXYACETONPHOSPHATE SYNTHASE (EC 2.5.1.26) (ALKYL-DHAP
DE SYNTHASE) (ALKYLGLYCERONE-PHOSPHATE SYNTHASE).
GN CG10253.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_Taxid=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burlingame K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheefer F., Shen H.,
RA Shue B.C., Slden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000)
CC -1- CATALYTIC ACTIVITY: 1-ACYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN
CC ALCOHOL = 1-ALKYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN ACID ANION.
CC -1- COFACTOR: FAD (BY SIMILARITY).

CC -1- PATHWAY: ETHER LIPID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FAD-BINDING OXIDOREDUCTASE/TRANSFERASE
CC FAMILY 4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF003812; AAF58181.1;
CC FlyBase: FBgn0035983; CG10253.
CC InterPro: IPR001575; Oxid_FAD_bind.
CC Pfam: PF01565; FAD_binding_4; 1.
CC Lipid synthesis; Transferase; Flavoprotein; FAD; Peroxisome.
CC SITE 629 631 MICROBODY TARGETING SIGNAL (POTENTIAL).
CC SITE 629 631
CC SEQUENCE 631 AA: 70821 MW: AADD6FDDA6BFB862 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 631;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RECS 5
DB 299 RECS 303

RESULT 11
BCCA_MYCTU STANDARD: PRT: 654 AA.
ID BCCA_MYCTU
AC P46401:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACETYL-/PROPYL-/COENZYME A CARBOXYLASE ALPHA CHAIN [INCLUDES: BIOTIN
DE CARBOXYLASE (EC 6.3.4.14); BIOTIN CARBOXYL CARRIER PROTEIN (BCCP)].
GN ACCA1 OR BCCA OR RV2501C OR MT2576 OR MTC107A7.07C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid=1773;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=50410;
RX MEDLINE=94222829; PubMed=7909542;
RA Norman E., de Smet K.A.L., Stoker N.G., Ratledge C., Wheeler P.R.,
RA Dale J.W.;
RT "Lipid synthesis in mycobacteria: characterization of the biotin
RT carboxyl carrier protein genes from Mycobacterium leprae and M.
RT tuberculosis."
RL J. Bacteriol. 176:2525-2531(1994).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekela F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
```


RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Blahut W.;
 RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and
 RT laboratory strains.";
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN CARRIES TWO FUNCTIONS: BIOTIN CARBOXYL
 CC CARRIER PROTEIN AND BIOTIN CARBOXYLTRANSFERASE.
 CC -1- CATALYTIC ACTIVITY: ATP + BIOTIN-CARBOXYL-CARRIER PROTEIN + CO(2)
 CC = ADP + OXTHIOPHOSPHATE + CARBOXYBIOTIN-CARBOXYL-CARRIER PROTEIN.
 CC -1- COFACTOR: BIOTIN.
 CC -1- PATHWAY: FIRST STEP IN LONG-CHAIN FATTY ACID SYNTHESIS.
 CC -1- SUBUNIT: MULTIMER COMPRISED OF 2 DIFFERENT SUBUNITS, THE LARGER
 CC ONE (63/64 KDA) HAS BIOTIN CARBOXYLASE AND BIOTIN CARRIER
 CC FUNCTIONS, WHILE THE SMALLER SUBUNIT POSSESSES CARBOXYLTRANSFERASE
 CC AND SUBSTRATE BINDING ACTIVITY.
 CC -1- SIMILARITY: TO OTHER BIOTIN-DEPENDENT ENZYMES AND CARBAMOYL-
 CC PHOSPHATE SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z19549; CAA79609.1; -
 DR EMBL: Z95556; CAB08919.1; -
 DR EMBL: AE007094; AAK46880.1; -
 DR HSSP: P24182; 1BNC.
 DR TIGR: M2576; -
 DR Tuberculin: RV2501c; -
 DR Interpro: IPR001882; Biotin.
 DR Interpro: IPR000089; Biotin_lipoyl.
 DR Interpro: IPR000901; Cpsase.
 DR Pfam: PF00364; biotin_lipoyl; 1.
 DR Pfam: PF00289; Cpsase_L_chain; 1.
 DR PROSITE: PS00188; BIOTIN; 1.
 DR PROSITE: PS00866; CPSASE_1; 1.
 DR PROSITE: PS00867; CPSASE_2; 1.
 KW Fatty acid biosynthesis; Ligase; Biotin; Multifunctional enzyme;
 KW ATP-binding; Complete proteome.
 FT DOMAIN 1 437 BIOTIN CARBOXYLASE.
 FT DOMAIN 588 654 BIOTIN CARBOXYL CARRIER PROTEIN.
 FT NP_BIND 162 167 ATP (BY SIMILARITY).
 FT ACT_SITE 294 294 BY SIMILARITY.
 FT BINDING 620 620 BIOTIN (BY SIMILARITY).
 SQ SEQUENCE 654 AA; 70592 MW; FA0A1A46432CABF CRC64;

Query Match 83.3%; Score 5; DB 1; Length 654;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
 DB 226 ERECS 230

RESULT 12
 PCCA_HUMAN STANDARD; PRT; 703 AA.
 AC P05165; O15979;
 DT 13-AUG-1987 (rel. 05, Created)
 DT 15-JUL-1999 (rel. 38, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE PROPYLONYL-COA CARBOXYLASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR
 DE (EC 6.4.1.3) (PCCA5 ALPHA SUBUNIT) (PROPIONYL-COA:CARBON DIOXIDE
 DE LIGASE ALPHA SUBUNIT).
 GN PCCA.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 RA Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89296507; PubMed-2740237;
 RA Lamouneh A.-M., Mahuran D.J., Gravel R.A.;
 RT "Human mitochondrial propionyl-CoA carboxylase: localization of the
 RT N-terminus of the pro- and mature alpha chains in the deduced
 RT primary sequence of a full-length cDNA.";
 RL Nucleic Acids Res. 17:4396-4396(1989).
 RN (2)
 RP REVISIONS..
 RA Gravel R.;
 RL Submitted (Apr-1993) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 344-536 FROM N.A.
 RX MEDLINE-86259695; PubMed-3460076;
 RA Lamouneh A.-M., Barankiewicz T.J., Willard H.F., Mahuran D.J.,
 RA Quan F., Gravel R.A.;
 RT "Isolation of cDNA clones coding for the alpha and beta chains of
 RT human propionyl-CoA carboxylase: Chromosomal assignments and DNA
 RT polymorphisms associated with PCCA and PCGB genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4864-4868(1986).
 RN (4)
 RP SEQUENCE OF 339-367 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-93167265; PubMed-8434582;
 RA Stankevics J., Ledley F.D.;
 RT "Cloning of functional alpha propionyl CoA carboxylase and correction
 RT of enzyme deficiency in pcca fibroblasts.";
 RL Am. J. Hum. Genet. 52:144-151(1993).
 RN (5)
 RP SEQUENCE OF 608-703 FROM N.A.
 RX MEDLINE-87212051; PubMed-355348;
 RA Lamouneh A.-M., Quan F., Gravel R.A.;
 RT "Sequence homology around the biotin-binding site of human
 RT propionyl-CoA carboxylase and pyruvate carboxylase.";
 RL Arch. Biochem. Biophys. 254:631-636(1987).
 RN (6)
 RP REVIEW ON PA VARIANTS.
 RX MEDLINE-99433966; PubMed-10502773;
 RA Ugarte M., Perez-Cerda C., Rodriguez-Romero P., Desviat L.R., Perez B.,
 RA Richard E., Muro S., Campeau E., Onura T., Gravel R.A.;
 RT "Overview of mutations in the PCCA and PCGB genes causing propionic
 RT acidemia.";
 RL Hum. Mutat. 14:275-282(1999).
 RN (7)
 RP VARIANTS PA-1 P-50; K-204; G-343; V-354; R-643 AND C-687 DEL.
 RX MEDLINE-99263311; PubMed-10329019;
 RA Campeau E., Dupuis L., Leon-Del-Rio A., Gravel R.;
 RT "Coding sequence mutations in the alpha subunit of propionyl-CoA
 RT carboxylase in patients with propionic acidemia.";
 RL Mol. Genet. Metab. 67:11-22(1999).
 CC -1- CATALYTIC ACTIVITY: ATP + PROPANOL-COA + HCO(3)(-) = ADP +
 CC ORTHOPHOSPHATE + (S)-METHYLMALONYL-COA.
 CC -1- COFACTOR: BIOTIN.
 CC -1- PATHWAY: KEY ENZYME IN THE CATABOLIC PATHWAY OF ODD-CHAIN
 CC FATTY ACIDS, ISOLEUCINE, THREONINE, METHIONINE, AND VALINE.
 CC -1- SUBUNIT: PROBABLY A DODECAMER COMPOSED OF SIX BIOTIN-CONTAINING
 CC ALPHA SUBUNITS AND SIX BETA SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1- DISEASE: DEFECTS IN PCCA ARE THE CAUSE OF PROPIONIC ACIDEMIA, TYPE
 CC I (PA-1), A LIFE-THREATENING DISEASE CHARACTERIZED BY EPISODIC

CC VOMITING, LETHARGY AND KETOSIS, NEUTROPENIA, PERIODIC
 CC THROMBOCYTOPENIA, HYPOGAMAGLOBULINEMIA, DEVELOPMENTAL
 CC RETARDATION, AND INTOLERANCE TO PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X14608; CAA32763.1; -
 DR EMBL; M13572; AAA60035.1; -
 DR EMBL; S55656; AAB25345.1; -
 DR EMBL; M26121; AAA36424.1; -
 DR PIR; A24151; A24151.
 DR PIR; A27883; A27883.
 DR PIR; S04613; S04613.
 DR HSP; P24182; IBNC.
 DR MIM; 232000; -
 DR InterPro: IPR001882; Biotin.
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR Pfam: PF00364; biotin_lipoyl; 1.
 DR Pfam: PF00289; CPSase_L_chain; 1.
 DR PROSITE; PS00188; BIOTIN; 1.
 DR PROSITE; PS00866; CPSASE_1; 1.
 DR PROSITE; PS00867; CPSASE_2; 1.
 DR PROSITE; PS00867; CPSASE_2; 1.
 KM Mitochondrion; Ligase; Biotin; ATP-binding; Transit peptide;
 KM Disease mutation; Polymorphism.
 FT TRANSIT 1 20
 FT CHAIN 21 703
 FT NP_BIND 198 203
 FT ACT_SITE 328 328
 FT BINDING 669 669
 FT VARIANT 50 50
 FT VARIANT 52 52
 FT VARIANT 113 113
 FT VARIANT 139 139
 FT VARIANT 204 204
 FT VARIANT 272 272
 FT VARIANT 343 343
 FT VARIANT 348 348
 FT VARIANT 354 354
 FT VARIANT 374 374
 FT VARIANT 398 398
 FT VARIANT 450 450
 FT VARIANT 534 534
 FT VARIANT 606 606
 FT VARIANT 643 643
 FT VARIANT 687 687
 FT VARIANT 344 366
 FT CONFLICT 533 533
 FT CONFLICT 654 654
 FT SEQUENCE 703 AA; 77353 MW; 0ABA6470BBE74ECE CRC64;

Query Match 83.3%; Score 5; DB 1; Length 703;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ERECS 5
 Db 262 ERECS 266
 RESULT 13
 ID PCQA_RAT STANDARD; PRT; 704 AA.
 AC P14882;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROPYLONYL-COA CARBOXYLASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR
 DE (EC 6.4.1.3) (PCCASE ALPHA SUBUNIT) (PROPYLONYL-COA:CARBON DIOXIDE
 DE LIGASE ALPHA SUBUNIT) (FRAGMENT).
 GN PCQA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89308706; PubMed=2745462;
 RA Browner M.F., Taroni F., Szul E., Rosenberg L.E.;
 RT "Sequence analysis, biogenesis, and mitochondrial import of the
 RT alpha-subunit of rat liver propionyl-CoA carboxylase.";
 RL J. Biol. Chem. 264:12680-12685(1989).
 RL [2]
 RP REVISIONS.
 RA Browner M.F., Taroni F., Szul E., Rosenberg L.E.;
 RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + PROPIONYL-COA + HCO(3)(-) - ADP +
 CC ORTHOPHOSPHATE + (S)-METHYLMALONYL-COA.
 CC -1- COFACTOR: BIOTIN.
 CC -1- PATHWAY: KEY ENZYME IN THE CATABOLIC PATHWAY OF ODD-CHAIN
 CC FATTY ACIDS. ISOLEUCINE, THREONINE, METHIONINE, AND VALINE.
 CC -1- SUBUNIT: PROBABLY A DODECAMER COMPOSED OF SIX BIOTIN-CONTAINING
 CC ALPHA SUBUNITS AND SIX BETA SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1- DISEASE: PROPIONIC ACIDEMIA DUE TO RECESSIVELY INHERITED
 CC DEFICIENCY OF PCCASE ACTIVITY OFTEN CAUSES LIFE-THREATENING
 CC KETOSIS AND ACIDOSIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M22631; AAA88512.1; ALT_SEQ.
 DR PIR; A34337; A34337.
 DR HSP; P24182; IBNC.
 DR InterPro: IPR001882; Biotin.
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR InterPro: IPR000901; CPSase.
 DR Pfam: PF00364; biotin_lipoyl; 1.
 DR Pfam: PF00289; CPSase_L_chain; 1.
 DR PROSITE; PS00188; BIOTIN; 1.
 DR PROSITE; PS00866; CPSASE_1; 1.
 DR PROSITE; PS00867; CPSASE_2; 1.
 KM Mitochondrion; Ligase; Biotin; ATP-binding; Transit peptide.
 FT NON_TER 1 1
 FT TRANSIT <1 21
 FT CHAIN 22 704
 FT NP_BIND 199 204
 FT MITOCHONDRION.
 FT PROPYLONYL-COA CARBOXYLASE ALPHA CHAIN.
 FT ATP (POTENTIAL).

FT ACT_SITE 329 329 BY SIMILARITY.
 FT BINDING 670 670 BIOTIN (BY SIMILARITY).
 SQ SEQUENCE 704 AA; 77711 MM; 36CEC52DF2D2A8A CRC64;

Query Match 83.3%; Score 5; DB 1; Length 704;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
 DB 263 ERECS 267

RESULT 14
 PIRK_DICDI STANDARD; PRT; 1093 AA.
 AC P54677;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHATIDYLINOSITOL 4-KINASE (EC 2.7.1.67) (PI4-KINASE) (PTDINS-4-KINASE) (PI4K-ALPHA).
 GN PIRK OR PIRK4.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=AX3;
 RX MEDLINE=96009592; PubMed=7565716;
 RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
 RT "A phosphatidylinositol (PI) kinase gene family in Dictyostellium
 discoideum: biological roles of putative mammalian p110 and yeast
 Vps34p PI 3-kinase homologs during growth and development.";
 RL Mol. Cell. Biol. 15:5645-5656(1995).
 CC -1- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST
 COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER
 INOSITOL-1,4,5,-TRISPHOSPHATE.
 CC -1- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL -> ADP +
 1-PHOSPHATIDYL-ID-MYO-INOSITOL 4-PHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U23479; AAA85725.1; -;
 DR DictyDb: DD01102; PIRK.
 DR InterPro: IPR000403; PI3_P14_Kinase.
 DR Pfam: PF00454; PI3_P14_Kinase; 1.
 DR SMART: SM00146; PI3Kc.1
 DR PROSITE: PS00915; PI3_4_KINASE_1; 1.
 DR PROSITE: PS00916; PI3_4_KINASE_2; 1.
 DR PROSITE: PS50290; PI3_4_KINASE_3; 1.
 KW transferase; Kinase; Multigene family.
 FT DOMAIN 146 156 ASP-RICH.
 FT DOMAIN 184 200 POLY-ASN.
 FT DOMAIN 203 206 POLY-ASN.
 FT DOMAIN 227 233 POLY-ASN.
 FT DOMAIN 233 315 POLY-ASN.
 FT DOMAIN 441 445 POLY-ASP.
 FT DOMAIN 455 466 POLY-THR.
 FT DOMAIN 494 501 POLY-GLY.
 FT DOMAIN 690 696 POLY-THR.
 FT DOMAIN 751 754 POLY-GLN.
 FT DOMAIN 761 772 POLY-THR.
 FT DOMAIN 775 785 POLY-THR.
 FT DOMAIN 833 1093 PI3K/PI4K.

SQ SEQUENCE 1093 AA; 123017 MM; 8762BC78355AA635 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 1093;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
 DB 1058 ERECS 1062

RESULT 15
 CFAH_MOUSE
 ID CFAH_MOUSE STANDARD; PRT; 1234 AA.
 AC P06909;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).
 GN CFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP MEDLINE=8623353; PubMed=2940596;
 RA Kristensen T., Tack B.F.;
 RT "Murine protein H is comprised of 20 repeating units, 61 amino acids
 in length.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
 RN [2]
 RC SEQUENCE OF 1-19 FROM N.A.
 RP STRAIN=BALB/C;
 RX MEDLINE=90148935; PubMed=2533512;
 RA Munoz-Canoves P., Tack B.F., Vlk D.P.;
 RT "Analysis of complement factor H mRNA expression: dexamethasone and
 IFN-gamma increase the level of H in L cells.";
 RL Biochemistry 28:9891-9897(1989).
 RN [3]
 RC SEQUENCE OF 1-18 FROM N.A.
 RX MEDLINE=90111033; PubMed=2136885;
 RA Matsume-Sakai S., Nonaka M., Harada Y.N., Shreffler D.C.,
 RA Moriwaki K.;
 RT "Demonstration of an unusual allelic variation of mouse factor H by
 the complete cDNA sequence of the H.2 allotype.";
 RL J. Immunol. 144:358-362(1990).
 CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COPROCTOR IN THE INACTIVATION OF
 C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
 C3BB COMPLEX (C3 CONVERTASE) AND THE (C3b)NB COMPLEX (C5
 CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
 CC -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
 MICE.
 CC -----
 CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M12660; AAA37759.1; -;
 DR EMBL: J02891; AAA37795.1; -;
 DR EMBL: M31979; AAA37762.1; -;
 DR PIR: A26154; NEMSH.
 DR HSSP: P06603; IHE1.
 DR MGD: MGI:88385; CFH.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00084; sush1; 20.
 DR SMART: SM00032; CCP; 20.

Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;									
KW	Signal	1	18	BY SIMILARITY.					
FT	SIGNAL	19	1234	COMPLEMENT FACTOR H.					
FT	CHAIN	20	81	SUSHI 1.					
FT	DOMAIN	19	1234	SUSHI 1.					
FT	DOMAIN	145	142	SUSHI 2.					
FT	DOMAIN	145	206	SUSHI 3.					
FT	DOMAIN	209	263	SUSHI 4.					
FT	DOMAIN	266	321	SUSHI 5.					
FT	DOMAIN	324	386	SUSHI 6.					
FT	DOMAIN	388	443	SUSHI 7.					
FT	DOMAIN	447	506	SUSHI 8.					
FT	DOMAIN	508	565	SUSHI 9.					
FT	DOMAIN	568	623	SUSHI 10.					
FT	DOMAIN	628	684	SUSHI 11.					
FT	DOMAIN	689	744	SUSHI 12.					
FT	DOMAIN	751	803	SUSHI 13.					
FT	DOMAIN	807	862	SUSHI 14.					
FT	DOMAIN	866	932	SUSHI 15.					
FT	DOMAIN	935	990	SUSHI 16.					
FT	DOMAIN	993	1049	SUSHI 17.					
FT	DOMAIN	1052	1108	SUSHI 18.					
FT	DOMAIN	1113	1169	SUSHI 19.					
FT	DOMAIN	1171	1234	SUSHI 20.					
FT	DISULFID	21	66	BY SIMILARITY.					
FT	DISULFID	52	80	BY SIMILARITY.					
FT	DISULFID	85	129	BY SIMILARITY.					
FT	DISULFID	114	141	BY SIMILARITY.					
FT	DISULFID	146	192	BY SIMILARITY.					
FT	DISULFID	178	205	BY SIMILARITY.					
FT	DISULFID	210	251	BY SIMILARITY.					
FT	DISULFID	237	262	BY SIMILARITY.					
FT	DISULFID	267	309	BY SIMILARITY.					
FT	DISULFID	294	320	BY SIMILARITY.					
FT	DISULFID	325	374	BY SIMILARITY.					
FT	DISULFID	357	385	BY SIMILARITY.					
FT	DISULFID	389	431	BY SIMILARITY.					
FT	DISULFID	416	442	BY SIMILARITY.					
FT	DISULFID	448	494	BY SIMILARITY.					
FT	DISULFID	477	505	BY SIMILARITY.					
FT	DISULFID	509	553	BY SIMILARITY.					
FT	DISULFID	536	564	BY SIMILARITY.					
FT	DISULFID	569	610	BY SIMILARITY.					
FT	DISULFID	597	622	BY SIMILARITY.					
FT	DISULFID	629	672	BY SIMILARITY.					
FT	DISULFID	658	683	BY SIMILARITY.					
FT	DISULFID	690	732	BY SIMILARITY.					
FT	DISULFID	718	743	BY SIMILARITY.					
FT	DISULFID	752	791	BY SIMILARITY.					
FT	DISULFID	780	802	BY SIMILARITY.					
FT	DISULFID	808	850	BY SIMILARITY.					
FT	DISULFID	836	861	BY SIMILARITY.					
FT	DISULFID	867	920	BY SIMILARITY.					
FT	DISULFID	906	931	BY SIMILARITY.					
FT	DISULFID	936	978	BY SIMILARITY.					
FT	DISULFID	964	989	BY SIMILARITY.					
FT	DISULFID	994	1037	BY SIMILARITY.					
FT	DISULFID	1023	1048	BY SIMILARITY.					
FT	DISULFID	1053	1096	BY SIMILARITY.					
FT	DISULFID	1082	1107	BY SIMILARITY.					
FT	DISULFID	1114	1157	BY SIMILARITY.					
FT	DISULFID	1143	1168	BY SIMILARITY.					
FT	DISULFID	1172	1223	BY SIMILARITY.					
FT	DISULFID	1206	1233	BY SIMILARITY.					
FT	CARBOHYD	676	721	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	773	773	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	801	801	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	1030	1030	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	1061	1061	N-LINKED (GLCNAC. . .) (POTENTIAL).					
FT	CARBOHYD	1225	1225	N-LINKED (GLCNAC. . .) (POTENTIAL).					
SO	SEQUENCE	1234	AA; 139082	MM; C5AC02F341B957F7	CRC64;				

Query Match 83.3%; Score 5; DB 1; Length 1234;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ERECS 5
DB 566 ERECS 570

Search completed: February 26, 2002, 01:53:56
Job time: 425 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:51:47 ; Search time 45.36 Seconds
(without alignments)
10.076 Million cell updates/sec

Title: US-09-819-136-2_COPY_50_55

Perfect score: 6

Sequence: 1 ERECSR 6

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	83.3	128	2	D69880
2	5	83.3	193	2	C84405
3	5	83.3	196	2	T00702
4	5	83.3	216	2	A39926
5	5	83.3	216	2	A48343
6	5	83.3	267	2	T16677
7	5	83.3	270	2	D72334
8	5	83.3	286	2	A72583
9	5	83.3	290	2	S50050
10	5	83.3	329	2	T50117
11	5	83.3	337	2	A81101
12	5	83.3	337	2	S37799
13	5	83.3	370	2	A71143
14	5	83.3	444	2	C70444
15	5	83.3	456	1	KXBO
16	5	83.3	458	2	A83585
17	5	83.3	458	2	B86600
18	5	83.3	458	2	B72024
19	5	83.3	506	2	D69277
20	5	83.3	565	2	T47330
21	5	83.3	568	2	F71614
22	5	83.3	571	2	C71528
23	5	83.3	654	2	B55579
24	5	83.3	665	2	T02793
25	5	83.3	665	2	C71667
26	5	83.3	667	2	C70719
27	5	83.3	669	2	S65551
28	5	83.3	703	1	A27883
29	5	83.3	704	2	A34337

30	5	83.3	724	2	T16187	hypothetical prote
31	5	83.3	735	2	E69139	hypothetical prote
32	5	83.3	991	2	C96764	hypothetical prote
33	5	83.3	1093	2	T18275	1-phosphatidylinos
34	5	83.3	1222	2	T14805	hypothetical prote
35	5	83.3	1234	1	NBMSH	complement factor
36	5	83.3	1279	2	T18312	hypothetical prote
37	5	83.3	1323	2	H85202	hypothetical prote
38	5	83.3	1868	2	T23707	hypothetical prote
39	5	83.3	3161	2	T30342	protein HMP1 - ye
40	5	83.3	3163	2	T17440	zinc finger protei
41	4	66.7	16	2	D45193	hypothetical prote
42	4	66.7	30	2	H72312	hypothetical prote
43	4	66.7	51	2	F82001	hypothetical prote
44	4	66.7	59	2	F85759	hypothetical prote
45	4	66.7	65	2	G69100	hypothetical prote

ALIGNMENTS

RESULT 1
D69880
hypothetical protein y1gd - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: D69880
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Huilo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardinc
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestre, P.; Tognoni, A.; Tostato, V.; Uchiy
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zundstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: D69880
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-128 <KUN>
A:Cross-references: GB:599112; GB:AL009126; NID:q2633902; PIDN:CB13474.1; PID:q2633939
A:Experimental source: strain 168
C:Genetics:
A:Gene: y1gd
C:Superfamily: Bacillus subtilis hypothetical protein y1gd
Query Match 83.3%; Score 5; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ERECS 5
Db 33 ERECS 37
RESULT 2
C84405
hypothetical protein Vng2554h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84405
R:Ng, W.V.; Kennedy, S.P.; Mahiras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
J.; Lettner, B.; Keller, K.; Cruz, R.; Danson, M.D.; Hough, D.W.; Maddocks, D.G.; Je
Jung, K.H.; Alam, M.; Freltas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: C84405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-193 <STO>
A:Cross-references: GB:AE004437; NID:g10581944; PIDN:AG20607.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG2554H

Query Match 83.3%; Score 5; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
|||||
Db 172 ERECS 176

RESULT 3

hypothetical protein F25965_1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 05-Nov-1999
C:Accession: T00702
R:Ramerdin, J.E.; McCreedy, P.M.; Adamson, A.W.; Burkhardt-Schultz, K.; Garcia, E.; Kyle,
h, A.; Olsen, A.O.; Carrano, A.V.
Submitted to the EMBL Data Library, October 1997
A:Description: Sequence analysis of a 1Mb region in 19q13.1.
A:Reference number: Z14199
A:Accession: T00702
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-196 <LAMB>
A:Cross-references: EMBL:AC002398; NID:g2529398; PIDN:AAB81199.1; PID:g2477514
C:Genetics:
A:Map position: 19
A:Note: Intron positions not resolved (incomplete sequence)
A:Note: F25965_1

Query Match 83.3%; Score 5; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
|||||
Db 81 ERECS 85

RESULT 4

hypothetical protein 1 - chicken anemia virus
C:Species: chicken anemia virus, CAV
C:Date: 17-Jan-1992 #sequence_revision 17-Jan-1992 #text_change 08-Oct-1999
C:Accession: A39926
R:Notelborn, M.H.M.; de Boer, G.F.; van Roozelaar, D.J.; Karreman, C.; Kranenburg, O.; VC
J. Virol. 65, 3131-3139, 1991
A:Title: Characterization of cloned chicken anemia virus DNA that contains all elements
A:Reference number: A39926; MUID:91237831
A:Accession: A39926
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <NOT>
A:Cross-references: GB:M55918; NID:g3323250; PIDN:AAA91822.1; PID:g3323251
C:Comment: This virus is unclassified.

Query Match 83.3%; Score 5; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
|||||
Db 85 RECSR 89

RESULT 5

A48343
24K hypothetical protein - chicken anemia virus (isolate Cux-1)
C:Species: chicken anemia virus, CAV
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 08-Oct-1999
C:Accession: A48343
R:Meenan, B.M.; Todd, D.; Creelan, J.L.; Earle, J.A.; Hoey, E.M.; McNulty, M.S.
Arch. Virol. 124, 301-319, 1992
A:Title: Characterization of viral DNAs from cells infected with chicken anaemia agen
A:Reference number: A48343; MUID:92296898
A:Accession: A48343
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <MEP>
A:Cross-references: GB:M81223; NID:g3323254; PIDN:AAA42882.1; PID:g3323255
A:Note: sequence extracted from NCBI backbone (NCBIN:106168, NCBI:P:106169)

Query Match 83.3%; Score 5; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
|||||
Db 85 RECSR 89

RESULT 6

hypothetical protein R04A9.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000
C:Accession: T16677
R:Geisels, C.
Submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid R04A9.
A:Reference number: Z18558
A:Accession: T16677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-267 <GER>
A:Cross-references: EMBL:U41550; NID:g1118045; PID:g1118047; PIDN:AAA83285.1; CESP:R0
C:Genetics:
A:Gene: CESP:R04A9.3
A:Introns: 15/3; 44/3; 80/2; 136/3; 160/1; 197/1; 250/1
C:Superfamily: *Caenorhabditis elegans* hypothetical protein R04A9.3

Query Match 83.3%; Score 5; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
|||||
Db 248 ERECS 252

RESULT 7

D72334
sugar kinase, pikB family - *Thermotoga maritima* (strain MSB8)
C:Species: *Thermotoga maritima*
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: D72334
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: D72334
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <ARN>
A:Cross-references: GB:AE001747; GB:AE000512; NID:g4981304; PIDN:AND35877.1; PID:g498132
A:Experimental source: strain MS8
C:Genetics:
A:Gene: TM0795

Query Match 83.3%; Score 5; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
|||||
DB 161 ERECS 165

RESULT 8
A72583
hypothetical protein APE1941 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: A72583
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339
A:Accession: A72583
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <KAM>
A:Cross-references: DDBJ:AF000062; NID:g5105244; PIDN:BAA80950.1; PID:g5105637
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1941
C:Superfamily: Aeropyrum pernix hypothetical protein APE1941

Query Match 83.3%; Score 5; DB 2; Length 286;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
|||||
DB 195 RECSR 199

RESULT 9
S50050
hypothetical protein 1 - Proteus vulgaris
C:Species: Proteus vulgaris
C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 08-Oct-1999
C:Accession: S50050
R:Janosi, L.; Yonemitsu, H.; Hong, H.; Kajl, A.
J. Mol. Biol. 242, 45-61, 1994
A:Title: Molecular cloning and expression of a novel hydroxymethylcytosine-specific rest
A:Reference number: S50050; MUID:94358863
A:Accession: S50050
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <JAN>
A:Cross-references: EMBL:U03474; NID:g424116; PIDN:AAA21850.1; PID:g424117
A:Note: the authors translated the codon TTC for residue 71 as Glu
C:Genetics:
A:Start codon: GTG

Query Match 83.3%; Score 5; DB 2; Length 290;

Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERECS 5
|||||
DB 260 ERECS 264

RESULT 10
T50117
mos2 homolog [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 28-Jul-2000
C:Accession: T50117
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
Submitted to the EMBL Data Library, February 2000
A:Reference number: Z25039
A:Accession: T50117
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-329 <SEE>
A:Cross-references: EMBL:AL157734; PIDN:CAB75774.1; GSPDB:GN00066; SPDB:SPAC1834.06c
A:Experimental source: strain 972h(-); cosmid c1834
C:Genetics:
A:Gene: SPDB:SPAC1834.06c
A:Map position: 1
A:Insertions: 34/3; 185/3
C:Superfamily: Saccharomyces hypothetical protein YKL189w

Query Match 83.3%; Score 5; DB 2; Length 329;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RECSR 6
|||||
DB 144 RECSR 148

RESULT 11
AB1101
C-5 cytosine-specific DNA-methylase NMB1290 [imported] - Neisseria meningitidis (stra
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
C:Accession: AB1101; F81841
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rl, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignanl, V.; Pizze, M.
Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20157555
A:Accession: AB1101
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <TEY>
A:Cross-references: GB:AE002477; GB:AE002098; NID:g7226521; PIDN:AAE41666.1; PID:g7222
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: AB1775; MUID:2022556
A:Accession: F81841
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <PAR>
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84733.1; PID:g738
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMB1290; NMA1500
C:Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII

Query Match 83.3%; Score 5; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RECSR 6
DB 288 RECSR 292

RESULT 12
S37799

hypothetical protein YKL151c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein YKL606

C:Species: *Saccharomyces cerevisiae*

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999

C:Accession: S37799; S37981; S44576

R:Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, September 1993

A:Description: DNA sequencing of a 36.2 Kb fragment located between the FAS1 and LAP4 loci

A:Reference number: S37786

A:Accession: S37799

A:Molecule type: DNA

A:Residues: 1-337 <V>

A:Cross-references: EMBL:Z26877; NID:9407482; PIDN:CAAB1502.1; PID:9407496

R:Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, March 1994

A:Reference number: S37976

A:Accession: S37981

A:Molecule type: DNA

A:Residues: 1-337 <V>

A:Cross-references: EMBL:Z28151; NID:9486261; PIDN:CAAB1992.1; PID:9486262; MIPS:YKL151C

R:Vandenbol, M.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, March 1994

A:Reference number: S44563

A:Accession: S44576

A:Molecule type: DNA

A:Residues: 1-337 <V>

A:Cross-references: EMBL:Z26877; NID:9407482; PIDN:CAAB1502.1; PID:9407496

A:Experimental source: strain S288C

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1993

C:Genetics:

A:Map position: 11L

Query Match 83.3%; Score 5; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RECSR 6
DB 305 RECSR 309

RESULT 13
A71143

hypothetical protein PH0356 - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: A71143

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekita, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon

A:Reference number: A71000; MUID:99344137

A:Accession: A71143

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Cross-references: GB:AP000002; NID:93236129; PIDN:BA29430.1; PID:93256747
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0356
C:Superfamily: hypothetical protein M0532

Query Match 83.3%; Score 5; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ERECS 5
DB 288 ERECS 292

RESULT 14
C70444

biotin carboxylase - *Aquifex aeolicus*

C:Species: *Aquifex aeolicus*

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999

C:Accession: C70444

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A:Reference number: A70300; MUID:98196666

A:Accession: C70444

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-444 <AOP>

A:Cross-references: GB:AE000749; NID:92983975; PIDN:AC07527.1; PID:92983981; GB:AE00

A:Experimental source: strain VPS

C:Genetics:

A:Gene: accC1

C:Superfamily: biotin carboxylase; biotin carboxylase homology

F:4-44/Domain: biotin carboxylase homology <BCH>

Query Match 83.3%; Score 5; DB 2; Length 444;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ERECS 5
DB 225 ERECS 229

RESULT 15
KXBO

protein C (activated) (EC 3.4.21.69) precursor <Bovine (fragment)

N:Alternate names: autoprothrombin IIA; Plasma protein C

C:Species: *Bos primigenius taurus* (cattle)

C:Date: 30-Nov-1980 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999

C:Accession: A26250; A18385; A18386; A00928

R:Long, G.L.; Balagaje, R.M.; MacGillivray, R.T.A.

Proc. Natl. Acad. Sci. U.S.A. 81, 5653-5656, 1984

A:Title: Cloning and sequence of liver cDNA coding for bovine protein C.

A:Reference number: A26250; MUID:85014826

A:Accession: A26250

A:Molecule type: mRNA

A:Residues: 1-456 <LON>

R:Fernlund, P.; Stenflo, J.

J. Biol. Chem. 257, 12170-12179, 1982

A:Title: Amino acid sequence of the light chain of bovine protein C.

A:Reference number: A18385; MUID:83007325

A:Accession: A18385

A:Molecule type: protein

A:Residues: 40-194 <PER>

A:Note: 82-Lys was also found

R:Drakenberg, T.; Fernlund, P.; Roepstorff, P.; Stenflo, J.

Proc. Natl. Acad. Sci. U.S.A. 80, 1802-1806, 1983

A:/Title: beta-Hydroxyaspartic acid in vitamin K-dependent protein C.
 A:/Reference number: A19316; MUID:83169769
 A:/Contents: annotation; revision to residue 110
 R:/Stenflo, J.; Fernlund, P.
 J. Biol. Chem. 257, 12180-12190, 1982
 A:/Title: Amino acid sequence of the heavy chain of bovine protein C.
 A:/Reference number: A18386; MUID:83007326
 A:/Accession: A18386
 A:/Molecule type: protein
 A:/Residues: 197-454,'PV'<STF>
 R:/Esmon, N.L.; Debault, L.E.; Esmon, C.T.
 J. Biol. Chem. 258, 5548-5553, 1983
 A:/Title: Proteolytic formation and properties of gamma-carboxyglutamic acid-domainless F
 A:/Reference number: A37541; MUID:83213513
 A:/Contents: annotation; activation; calcium binding
 R:/Johnson, A.E.; Esmon, N.L.; Laue, T.W.; Esmon, C.T.
 J. Biol. Chem. 258, 5554-5560, 1983
 A:/Title: Structural changes required for activation of protein C are induced by Ca²⁺ but
 A:/Reference number: A37542; MUID:83213514
 A:/Contents: annotation; activation; calcium binding
 C:/Comment: Protein C is the zymogen of the vitamin K-dependent serine proteinase that re
 S:
 C:/Comment: Protein C is synthesized in the liver as a single chain precursor, which is c
 bin, which cleaves a tetradecapeptide from the amino end of the heavy chain; this react
 C:/Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with strc
 cognition of the thrombin-thrombomodulin complex.
 C:/Comment: The gamma-carboxyglutamic acid residues arise by a posttranslational, vitamin
 C:/Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
 C:/Keywords: anticoagulant; beta-hydroxyaspartic acid; blood coagulation; calcium binding
 F:/1-29/Domain: signal sequence (fragment) #status predicted <SIG>
 F:/24-83/Domain: Gla domain homology <GLA>
 F:/30-39/Domain: propeptide #status predicted <PRO>
 F:/40-184/Product: protein C light chain #status experimental <LCH>
 F:/98-128/Domain: EGF homology <EG1>
 F:/137-172/Domain: EGF homology <EG2>
 F:/197-456/Product: protein C heavy chain #status experimental <HCH>
 F:/197-210/Domain: activation peptide #status experimental <APT>
 F:/211-440/Domain: trypsin homology <TRY>
 F:/45,46,53,55,58,59,62,64,65,68,74/Modified site: gamma-carboxyglutamic acid (Glu) #stat
 F:/110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
 F:/119-128,137-148,144-157,159-172,180-318,237-253,368-382,393-421/Disulfide bonds: #stat
 F:/136,289,350/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:/252,298,397/Active site: His, Asp, Ser #status predicted
 F:/366/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 83.3%; Score 5; DB 1; Length 456;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ERECS 5
 11111
 Db 53 ERECS 57

Search completed: February 26, 2002, 01:51:48
 Job time: 457 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:53:54 ; Search time 24.25 Seconds
(without alignments)
9.072 Million cell updates/sec

Title: US-09-819-136-2_COPY_283_288
Perfect score: 6
Sequence: 1 REPAR 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36654827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5	83.3	124	1 PYP_RHOCA	O69138 rhododactyl
2	5	83.3	157	2 RL12_CHURE	P50884 chlamydomon
3	5	83.3	165	1 RL12_SCHRO	P75000 schizosacch
4	5	83.3	216	1 GSPJ_ERMCA	P31589 ervinia car
5	5	83.3	267	1 COLI_PIG	P01192 s corticotr
6	5	83.3	277	1 DKG_HUMAN	Q16854 homo sapien
7	5	83.3	305	1 Y452_BUCAI	P57527 buchnera ap
8	5	83.3	455	1 GAAL_CHICK	P19150 gallus gall
9	5	83.3	455	1 GAAL_MOUSE	P18504 mus musculu
10	5	83.3	456	1 GAAL_BOVIN	P08219 bos taurus
11	5	83.3	456	1 GAAL_HUMAN	P14867 homo sapien
12	5	83.3	468	1 SELA_PSEAE	Q9HVO1 pseudomonas
13	5	83.3	567	1 PRO1_LEIEN	P13865 leiishmania
14	5	83.3	570	1 FLIF_RHOSH	Q53151 rhododactyl
15	5	83.3	599	1 RF2P_DROME	P14199 drosophila
16	5	83.3	692	1 AMO_PICAN	P12807 pichia angu
17	5	83.3	742	1 PKWA_THECU	P43695 thermomonas
18	5	83.3	743	1 TPE3_HUMAN	P19532 homo sapien
19	5	83.3	808	1 PLD_PIMR	O04883 plimnelli
20	5	83.3	808	1 PLD_TOBAC	P93400 nicotiana t
21	5	83.3	1275	1 RFBG_MYXXX	O50864 mycoccocus
22	5	83.3	1380	1 SPI4_YEAST	P36126 saccharomyc
23	5	83.3	5032	1 RYNR_HUMAN	P21817 homo sapien
24	5	83.3	5035	1 RYNR_PIG	P19660 sus scrofa
25	5	83.3	5037	1 RYNR_RABIT	P17116 oryctolagus
26	4	66.7	50	1 AROQ_AMEYE	P46380 amycolatops
27	4	66.7	50	1 BAL3_SCHRO	Q02594 schizosphyll
28	4	66.7	54	1 VHAL_ECOLI	P42625 escherichia
29	4	66.7	55	1 RLX_HALMA	P14125 halococcus
30	4	66.7	66	1 KED2_ECOLI	P13967 escherichia
31	4	66.7	67	1 LEGX_HUMAN	P05163 homo sapien
32	4	66.7	68	1 S61G_CAEEL	Q1967 caenorhabdit
33	4	66.7	69	1 H6_ONCMY	P02315 oncorhynchus

34	4	66.7	73	1 PHD_BPPI	O06253 bacterioph
35	4	66.7	83	1 YC2_PARDE	P29969 paracoccus
36	4	66.7	86	1 Y425_TREPA	O83440 treponema p
37	4	66.7	97	1 XRA_CRIGR	O64029 cricetus
38	4	66.7	99	1 Y107_MYCTU	O50718 mycobacteri
39	4	66.7	100	1 PIML_HUMAN	O15428 homo sapien
40	4	66.7	102	1 MGN_ORYZA	P49030 oryza sativ
41	4	66.7	105	1 LEGX_MOUSE	P11946 mus musculu
42	4	66.7	111	1 YIFE_HAEIN	P31811 haemophilus
43	4	66.7	111	1 YIS9_YEAST	P40567 saccharomyc
44	4	66.7	120	1 TBFL_TUBBO	P80708 tuber borch
45	4	66.7	120	1 Y025_MYCTU	P71595 mycobacteri

ALIGNMENTS

RESULT 1	
PYP_RHOCA	STANDARD; PRT; 124 AA.
ID PYP_RHOCA	
AC O69138;	
DT 20-AUG-2001 (Rel. 40; Created)	
DT 20-AUG-2001 (Rel. 40; Last sequence update)	
DT 20-AUG-2001 (Rel. 40; Last annotation update)	
DE PHOTOACTIVE YELLOW PROTEIN (PYP).	
GN PYP.	
OS Rhododactyl capsulatus (Rhodopsudomonas capsulata).	
OC Bacteria; Proteobacteria; alpha subdivision; Rhododactyl group;	
OC Rhododactyl.	
OX NCBI_TaxID=1061;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=SB1003 / ST LOUIS;	
RA Jiang Z., Bauer E.C.;	
RT "Genetic characterization of photoactive yellow protein from Rhododactyl capsulatus";	
RL Submitted (MAY-1998) to the EMBL/Genbank/DDJ databases.	
CC - FUNCTION: THIS PHOTOACTIVE PROTEIN IS A PHOTORECEPTOR WITH KINETICS SIMILAR TO THAT OF RHODOPSIN (BY SIMILARITY).	
CC - PFM: COVALENTLY BIND A 4-HYDROXYCINNAMYL (P-COUMARIC ACID) CHROMOPHORE VIA A THIOESTER LINKAGE (BY SIMILARITY).	
CC - SIMILARITY: BELONGS TO THE XANTHOPSYN FAMILY.	
CC - SIMILARITY: CONTAINS 1 PAS DOMAIN.	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@sdb-sdb.ch).	
CC	
DR EMBL; AF064095; AAC17427.1; -	
DR HSSP; P16113; 2PYR.	
DR InterPro; IPR000104; PAS.	
DR SMART; SM00091; PAS; 1.	
KW Photoceptor; Chromophore.	
FT DOMAIN 22	
FT BINDING 68	
SO SEQUENCE 124 AA; 13950 MW; FDB4858D396BBA51 CRC64;	
Query Match	83.3%; Score 5; DB 1; Length 124;
Best Local Similarity	100.0%; Pred. No. 9.6;
Matches 5; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAR 5	
DB 17 REPAR 21	
RESULT 2	
RL12_CHURE	

```

ID  RL12_CHLRE  STANDARD;  PRT;  157 AA.
AC  P50884;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  60S RIBOSOMAL PROTEIN L12 (FRAGMENT).
GN  RPL12.
OS  Chlamydomonas reinhardtii.
OC  Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC  Chlamydomonadales; Chlamydomonadales.
OX  NCBI_TaxID=3055;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=7781;
RA  Walter F.;
RL  Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 26S RIBOSOMAL RNA (BY
CC  SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE L11P FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X95314; CAA64626.1; -
DR  InterPro; IPR000911; Ribosomal_L11.
DR  Pfam; PF00298; Ribosomal_L11; 1.
DR  ProDom; PD001367; Ribosomal_L11; 1.
DR  PROSITE; PS00359; RIBOSOMAL_L11; 1.
KW  Ribosomal protein.
FT  NON_TER 1
SQ  SEQUENCE 157 AA; 16633 MW; 884C33677B30B8C1 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EPARD 6
Db 79 EPARD 83

RESULT 3
RL12_SCHPO
ID  RL12_SCHPO  STANDARD;  PRT;  165 AA.
AC  Q75000;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  60S RIBOSOMAL PROTEIN L12.
GN  (RPL12A OR SPOC3H12.04C) AND (RPL12B OR SPOC16C4.13C).
OS  Schizosaccharomyces pombe (Fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomycetes.
OX  NCBI_TaxID=4896;
RN  [1]
RP  SEQUENCE FROM N.A. (RPL12A).
RC  STRAIN=972;
RA  Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL  Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A. (RPL12B).
RC  STRAIN=972;
RA  Purrelle B., Goffeau A., Wood V., Rajandream M.A., Barrell B.G.;
RL  Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 26S RIBOSOMAL RNA (BY
CC  SIMILARITY).

```

```

CC  -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L12 IN S. POMBE.
CC  -1- SIMILARITY: BELONGS TO THE L11P FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AL031824; CAA21221.1; -
DR  EMBL; AL031535; CAA20752.1; -
DR  InterPro; IPR000911; Ribosomal_L11.
DR  Pfam; PF00298; Ribosomal_L11; 1.
DR  ProDom; PD001367; Ribosomal_L11; 1.
DR  PROSITE; PS00359; RIBOSOMAL_L11; 1.
KW  Ribosomal protein; RNA-binding; Multigene family.
SQ  SEQUENCE 165 AA; 17666 MW; FC777B46CD08F005 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EPARD 6
Db 87 EPARD 91

RESULT 4
GSPJ_ERWCA
ID  GSPJ_ERWCA  STANDARD;  PRT;  216 AA.
AC  P31589;
DT  01-JUL-1993 (Rel. 26, Created)
DT  01-JUL-1993 (Rel. 26, Last sequence update)
DT  01-JUL-1993 (Rel. 26, Last annotation update)
DE  GENERAL SECRETION PATHWAY PROTEIN J PRECURSOR (PECTIC ENZYMES
DE  SECRETION PROTEIN OUTJ).
GN  OUTJ.
OS  Erwinia carotovora.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Pectobacterium.
OX  NCBI_TaxID=554;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SCRI193;
RX  MEDLINE=93316842; PubMed=8326859;
RA  Reeves P.J., Whitcombe D., Wharam S., Gibson M., Allison G.,
RA  Bunce N., Barallion R., Douglas P., Mulholland V., Stevens S.,
RA  Walker S., Salmund G.P.C.;
RT  Molecular cloning and characterization of 13 out genes from Erwinia
RT  carotovora subspecies carotovora: genes encoding members of a general
RT  secretion pathway (GSP) widespread in gram-negative bacteria.;
RL  Mol. Microbiol. 8:443-456(1993).
CC  -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC  EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC  MULTIPLE PECTIC ENZYMES.
CC  -1- SIMILARITY: BELONGS TO THE PULJ/OUTJ/XPSJ/EXEJ/XCPW FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X70049; CAA49651.1; -
DR  PIR; S31753; S31753.
DR  PIR; S32864; S32864.
DR  InterPro; IPR003413; GSPJIL_IJ.
DR  InterPro; IPR001120; Prok_N_methyltn.

```

DR Pfam: PF02501; GSPRT_IJ: 1.
 DR PROSITE: PS00409; PROKAR_NTER_METHYL: 1.
 KW Transport; Methylation.
 FT PROPEP 1 25 BY SIMILARITY.
 FT CHAIN 26 216 GENERAL SECRETION PATHWAY PROTEIN J.
 FT MOD_RES 26 26 METHYLATION (BY SIMILARITY).
 SQ SEQUENCE 216 AA; 24850 MW; 96223BCA29471070 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 216;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 DB 15 REPAR 19

RESULT 5
 COLI_PIG STANDARD: PRT; 267 AA.
 ID COLI_PIG P01192; Q95246;
 AC 21-JUL-1986 (Rel. 01, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CORTICOTROPIN-LIPOPOTROPIN PRECURSOR (PRO-OPOMELANOCORTIN) (POMC)
 DE [CONTAINS: NPI; MELANOTROPIN GAMMA (GAMA-MSH); CORTICOTROPIN
 DE (ADRENOCORTICOTROPIC HORMONE) (ACTH); MELANOTROPIN ALPHA (ALPHA-MSH);
 DE CORTICOTROPIN-LIKE INTERMEDIARY PEPTIDE (CLIP); LIPOPOTROPIN BETA (BETA-
 DE LPH); LIPOPOTROPIN GAMMA (GAMA-LPH); MELANOTROPIN BETA (BETA-MSH);
 DE BETA-ENDORPHIN; MET-ENKEPHALIN].
 GN POMC.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86131687; PubMed=3753882;
 RA Gossard F.J., Chang A.C.Y., Cohen S.N.;
 RT "Sequence of the cDNA encoding porcine pro-opiomelanocortin.";
 RL Blochim. Biophys. Acta 866:68-74(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=84069823; PubMed=6196724;
 RA Boileau G., Barbeau C., Jeannotte L., Chretien M., Drouin J.;
 RT "Complete structure of the porcine pro-opiomelanocortin mRNA derived
 RT from the nucleotide sequence of cloned cDNA.";
 RL Nucleic Acids Res. 11:8063-8071(1983).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95046835; PubMed=7958386;
 RA Gen K., Hirai T., Kato T., Kato Y.;
 RT "Presence of the same transcript of pro-opiomelanocortin (POMC) genes
 RT in the porcine anterior and intermediate pituitary lobes.";
 RL Mol. Cell. Endocrinol. 103:101-108(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=84239667; PubMed=6547437;
 RA Oates E., Herbert E.;
 RT "5' sequence of porcine and rat pro-opiomelanocortin mRNA. One porcine
 RT and two rat forms.";
 RL J. Biol. Chem. 259:7421-7425(1984).
 RN [5]
 RP SEQUENCE OF 136-174.
 RA Shepherd R.G., Willson S.D., Howard K.S., Bell P.H., Davies D.S.,
 RA Davis S.B., Signer E.A., Shakespeare N.E.;
 RT "Studies with corticotropin, III. Determination of the structure of
 RT beta-corticotropin and its active degradation products.";
 RL J. Am. Chem. Soc. 78:5067-5076(1956).
 RN [6]
 RP REVISIONS TO 160 AND 165.
 RX MEDLINE=72114902; PubMed=4334191;

RA Runkler B., Sieber P., Rittel W., Zuber H.;
 RT "Revised amino acid sequences for porcine and human
 RT adrenocorticotrophic hormone.";
 RL Nature New Biol. 235:114-115(1972).
 RN [7]
 RP REVISIONS (CORTICOTROPIN).
 RA MEDLINE=74306590; PubMed=4369114;
 RA Graf L.;
 RT "Re-examination of the sequence of the C-terminal tryptic fragment
 RT from porcine adrenocorticotrophic hormone.";
 RL Acta Biochim. Biophys. Acad. Sci. Hung. 7:293-297(1972).
 RN [8]
 RP SEQUENCE OF 136-174.
 RA MEDLINE=91071194; PubMed=2174774;
 RA Voigt K., Stegmaler W., McGregor G.P., Roesch H., Seliger H.;
 RT "Isolation and full structural characterisation of six
 RT adrenocorticotropin-like peptides from porcine pituitary gland.
 RT Identification of three novel fragments of adrenocorticotropin and of
 RT two forms of a novel adrenocorticotropin-like peptide.";
 RL Eur. J. Biochem. 194:225-236(1990).
 RN [9]
 RP SEQUENCE OF 136-148.
 RA Harris J.I., Lerner A.B.;
 RT "Amino-acid sequence of the alpha-melanocyte-stimulating hormone.";
 RL Nature 179:1346-1347(1957).
 RN [10]
 RP SEQUENCE OF 177-267.
 RA MEDLINE=7111231; PubMed=5543613;
 RA Graf L., Barat E., Cseh G., Sajgo M.;
 RT "Amino acid sequence of porcine beta-lipotrophic hormone.";
 RL Blochim. Biophys. Acta 229:276-278(1971).
 RN [11]
 RP REVISIONS (LIPOPOTROPIN).
 RA Gilardieu C., Chretien M.;
 RT "Complete amino acid sequence of porcine beta-lipotrophic hormone
 RT (beta-LPH).";
 RL (In) Meienhofer J. (eds.);
 RL Chemistry and Biology of peptides, pp.609-611, Ann Arbor Sci. Pub.,
 RL Ann Arbor (1972).
 RN [12]
 RP REVISION TO 211.
 RA MEDLINE=73048217; PubMed=4673865;
 RA Pankov Y.A., Yudaev N.A.;
 RT "Complete amino acid sequence in the molecule of porcine beta-
 RT lipotropin.";
 RL Biokhimiia 37:991-1004(1972).
 RN [13]
 RP SEQUENCE OF 217-234.
 RA Harris J.I., Roos P.;
 RT "Amino-acid sequence of a melanophore-stimulating peptide.";
 RL Nature 178:90-90(1956).
 RN [14]
 RP SEQUENCE OF 217-234.
 RA Geschwind I.I., Li C.H., Barnafi L.;
 RT "The structure of the beta-melanocyte-stimulating hormone.";
 RL J. Am. Chem. Soc. 79:620-625(1957).
 RN [15]
 RP SEQUENCE OF 237-241.
 RA MEDLINE=76100762; PubMed=1207728;
 RA Hughes J., Smith T.W., Kosterlitz H.W., Fothergill L.A., Morgan B.A.,
 RA Morris H.R.;
 RT "Identification of two related pentapeptides from the brain with
 RT potent opiate agonist activity.";
 RL Nature 258:577-579(1975).
 RN [16]
 RP SEQUENCE OF 237-267.
 RA MEDLINE=77084500; PubMed=1007884;
 RA Graf L., Barat E., Patthy A.;
 RT "Isolation of a COOH-terminal beta-lipotropin fragment (residues
 RT 61-91) with morphine-like analgesic activity from porcine pituitary
 RT glands.";
 RL Acta Biochim. Biophys. Acad. Sci. Hung. 11:121-122(1976).
 CC -I- FUNCTION: ACTH STIMULATES THE ADRENAL GLANDS TO RELEASE CORTISOL.

CC -1- FUNCTION: MSH (MELANOCYTE-STIMULATING HORMONE) INCREASES THE
CC PIGMENTATION OF SKIN BY INCREASING MELANIN PRODUCTION IN
CC MELANOCYTES.
CC -1- FUNCTION: BETA-ENDORPHIN AND MET-ENKEPHALIN ARE ENDOGENOUS
CC OPIATES.
CC -1- TISSUE SPECIFICITY: ACTH AND MSH ARE PRODUCED BY THE PITUITARY
CC GLAND.
CC -1- SPECIFIC ENZYMOLOGICAL CLEAVAGES AT PAIRED BASIC RESIDUES YIELD
CC THE DIFFERENT ACTIVE PEPTIDES.
CC -1- SIMILARITY: BELONGS TO THE POMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03561; CA227248.1; -;
DR EMBL; X00135; CA24968.1; -;
DR EMBL; S73519; AAB32312.1; -;
DR EMBL; K01879; AAA31104.1; -;
DR PIR; A01454; CTPGP
DR InterPro: IPR001941; Melanocortin_ACTH.
DR Pfam; PF00976; ACTH_domain; 1.
DR PRINTS; PR00383; MELANOCORTIN.
KW Endorphin; Hormone; Cleavage on pair of basic residues; Amidation;
KW Glycoprotein; Signal.
FT SIGNAL 1 26
FT PEPTIDE 27 106 BY SIMILARITY.
FT PEPTIDE 77 87 NPY.
FT PEPTIDE 136 174 MELANOTROPIN GAMMA.
FT PEPTIDE 136 174 CORTICOTROPIN.
FT PEPTIDE 136 174 MELANOTROPIN ALPHA.
FT PEPTIDE 154 174 CORTICOTROPIN-LIKE INTERMEDIARY PEPTIDE.
FT PEPTIDE 177 267 LIPTROPIN BETA.
FT PEPTIDE 177 234 LIPTROPIN GAMMA.
FT PEPTIDE 217 234 MELANOTROPIN BETA.
FT PEPTIDE 237 267 BETA-ENDORPHIN.
FT PEPTIDE 237 241 MET-ENKEPHALIN.
FT MOD_RES 87 87 AMIDATION (G-88 PROVIDE AMIDE GROUP).
FT MOD_RES 148 148 AMIDATION (G-149 PROVIDE AMIDE GROUP).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 143 143 R -> T.
FT CONFLICT 6 6 G -> S (IN REF. 3 AND 4).
FT CONFLICT 15 15 T -> A (IN REF. 3 AND 4).
FT CONFLICT 23 23 G -> E (IN REF. 3 AND 4).
FT CONFLICT 49 49 A -> S (IN REF. 4).
SQ SEQUENCE 267 AA; 28894 MW; A6DB487A5032B648 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 267;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EPARD 6
Db 184 EPARD 188

RESULT 6
DCK_HUMAN STANDARD; PRT; 277 AA.
AC Q16854; Q16759; P78532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYGUANOSINE KINASE, MITOCHONDRIAL PRECURSOR (EC 2.7.1.113) (DCK).
GN DCUOK OR DCK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96293511; PubMed=8692979;
RA Johansson M., Karlsson A.;
RT "Cloning and expression of human deoxyguanosine kinase cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:7258-7262(1996).
RN [2]
RP SEQUENCE OF 18-277 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96314545; PubMed=8706825;
RA Wang L., Hellman U., Eriksson S.;
RT "Cloning and expression of human mitochondrial deoxyguanosine kinase
RL cDNA.";
RN [3]
RP SEQUENCE OF 1-47 FROM N.A.
RA Stegmann A.P.A., Mitchell B.S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL
CC DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY
CC EMPLOYED AS ANTIVITAL AND CHEMOTHERAPEUTIC AGENTS.
CC -1- CATALYTIC ACTIVITY: ATP + DEOXYGUANOSINE = ADP + DGMP.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4 AND 5;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS. HIGHEST EXPRESSION IN MUSCLE,
CC BRAIN, LIVER, AND LYMPHOID TISSUES.
CC -1- SIMILARITY: BELONGS TO THE DCK/DGK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41668; AAC50624.1; -;
DR EMBL; X97386; CAA6054.1; -;
DR EMBL; U81499; AAB39858.1; -;
DR MIM; 601465; -;
DR InterPro: IPR002624; DNK.
DR Pfam; PF01712; DNK.
KW Transferrase; kinase; ATP-binding; Mitochondrion; Transist peptide;
KW Alternative splicing.
FT TRANSIT 1 39
FT CHAIN 40 277
FT NP_BIND 45 52
FT VARSPPLIC 48 85
FT FT
FT VARSPPLIC 149 236
FT VARSPPLIC 47 47
FT CONFLICT 83 83 I -> IGLNLTQIGRAPLOET (IN ISOFORM 5).
FT CONFLICT 159 159 N -> T (IN REF. 2).
FT CONFLICT 212 212 G -> D (IN REF. 2).
FT CONFLICT 212 212 K -> E (IN REF. 2).
SQ SEQUENCE 277 AA; 32069 MW; D1F8AAC15224EE2 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 277;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAR 5
Db 101 REPAR 105

RESULT 7
Y452_BUCAI STANDARD; PRT; 305 AA.
ID Y452_BUCAI
AC P57527;

```

DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN BU452.
GN BU452.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
  symbiotic bacterium).
OS Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
  Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0003 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
  or send an email to license@isb-sdb.ch).
-----
DR EMBL: AP001119; BAB3150.1;
DR InterPro: IPR001880; UPF0003.
DR Pfam: PF00924; UPF0003.1.
DR PROSITE: PS01246; UPF0003.1.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 89 109 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
SQ SEQUENCE 305 AA; 34129 MW; 6F99B52D3EB9AA56 CRC64;
-----
Query Match 83.3%; Score 5; DB 1; Length 305;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAR 5
DB 204 REPAR 208
-----
RESULT 8
ID GAAL_CHICK STANDARD; PRT; 455 AA.
AC P19150;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-1 SUBUNIT PRECURSOR (GABA(A)
  RECEPTOR).
GN GABRA1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
  OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COMMERCIAL RHODE ISLAND CROSS; TISSUE=Brain;
RX MEDLINE=91245932; PubMed=1710013;
RA Bateson A.N., Harvey R.J., Wilsden W., Glencorse T.A., Hicks A.A.,
  Hunt S.P., Barnard E.A., Darlison M.G.;
RT "The chicken GABA receptor alpha 1 subunit: cDNA sequence and
  localization of the corresponding mRNA."
RL Brain Res. Mol. Brain Res. 9:333-339(1991).

```

```

CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE
CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE
CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE
  CHANNEL.
CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)
  CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
  or send an email to license@isb-sdb.ch).
-----
DR EMBL: X54244; CAA38148.1;
DR PIR: S11749; CHCHAL.
DR PIR: A60033; A60033.
DR InterPro: IPR001173; Neur_Channel.
DR Pfam: PF00065; neur_chan. 2.
DR PRINTS: PR00252; NRIONCHANNEL.
DR PRINTS: PR00253; GABARECEPTR.
DR PRINTS: PR01079; GABARALPHA.
DR PROSITE: PS00236; NEUOTR_ION_CHANNEL. 1.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; signal;
  KW Transmembrane; Multigene family.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 455 GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-1
  FT SUBUNIT EXTRACELLULAR (PROBABLE).
FT TRANSMEM 252 273 PROBABLE.
FT TRANSMEM 279 300 PROBABLE.
FT TRANSMEM 313 334 PROBABLE.
FT DOMAIN 335 420 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 421 441 PROBABLE.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFD 166 180 BY SIMILARITY.
SQ SEQUENCE 455 AA; 51842 MW; 23AD620C56762EBD CRC64;
-----
Query Match 83.3%; Score 5; DB 1; Length 455;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 REPAR 5
DB 200 REPAR 204
-----
RESULT 9
ID GAAL_MOUSE STANDARD; PRT; 455 AA.
AC P18504;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-1 SUBUNIT PRECURSOR (GABA(A)
  RECEPTOR).
GN GABRA1 OR GABRA-1.
OS Mus musculus (Mouse), and
  OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat;
RX MEDLINE=89211387; PubMed=2540033;
RA Lolait S.J., O'Carroll A.-M., Kusano K., Muller J.-M.,
  Brownstein M.J., Mahan L.C.;

```

RT "Cloning and expression of a novel rat GABA_A receptor.";
 RL FEBS Lett. 246:145-148(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Rat;
 RX MEDLINE=90380416; PubMed=2561977;
 RA Kreschatskiy M., MacLennan A.J., Chiang M.Y., Xu W., Jackson M.B.,
 Brechea N., Sterini C., Olsen R.W., Tobin A.J.;
 RT "A novel alpha subunit in rat brain GABA_A receptors.";
 RL Neuron 3:745-753(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Rat; TISSUE-Brain;
 RX MEDLINE=92111149; PubMed=1966765;
 RA Seeburg P.H., Wisden W., Verdoorn T., Pritchett D., Werner P.,
 Herb A., Lueddens H., Sprengel R., Sakmann B.;
 RT "The GABA_A receptor family: molecular and functional diversity.";
 RL Cold Spring Harb. Symp. Quant. Biol. 55:29-40(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Mouse; STRAIN=C57BL/6J, AND DBA/2J; TISSUE=Brain;
 RX MEDLINE=93002190; PubMed=1356407;
 RA Wang J.B., Kofuji P., Fernando J.C., Moss S.J., Huganir R.L.,
 Burt D.R.;
 RT "The alpha 1, alpha 2, and alpha 3 subunits of GABA_A receptors:
 RT comparison in seizure-prone and -resistant mice and during
 RT development.";
 RL J. Mol. Neurosci. 3:177-184(1992).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Mouse;
 RX MEDLINE=91169549; PubMed=1848528;
 RA Keir W.J., Kozak C.A., Chakraborti A., Deitrich R.A., Sikeja J.M.;
 RT "The cDNA sequence and chromosomal location of the murine GABA_A alpha
 RT 1 receptor gene.";
 RL Genomics 9:390-395(1991).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Mouse; STRAIN-BALB/C;
 RA Kato K.;
 RT "A collection of cDNA clones with specific expression patterns in
 RT mouse brain.";
 RL Eur. J. Neurosci. 2:704-711(1991).
 RN [7]
 RP MUTAGENESIS OF PHE-91.
 RX MEDLINE=92289664; PubMed=1376242;
 RA Sigel E., Baur R., Kellenberger S., Malherbe P.;
 RT "Point mutations affecting antagonist affinity and agonist dependent
 RT gating of GABA_A receptor channels.";
 RL EMBO J. 11:2017-2023(1992).
 CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE
 CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE
 CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE
 CC CHANNEL.
 CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)
 CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: I08490; AAC42029.1; -
 DR EMBL: M86566; AAB59634.1; -
 DR EMBL: M63436; AAB57654.1; -
 DR EMBL: X61430; CAA43672.1; -
 DR PIR: A39062; A39062.
 DR PIR: S03889; S03889.

DR PIR: J00158; J00158.
 DR PIR: S37710; S37710.
 DR MGD: MGI:95613; Gabral.
 DR InterPro: IPR001175; Neur_channel.
 DR Pfam: Pf00065; neur_chan. 2.
 DR PRINTS: PR00252; NRIONCHANNEL.
 DR PRINTS: PR00253; GABAARECEPT.
 DR PRINTS: PR01079; GABAARALPHA.
 DR PROSITE: PS00236; NEUROTR_ION_CHANNEL; 1.
 KW Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
 KW Transmembrane; Multigene family.
 FT SIGNAL 1 27
 FT CHAIN 28 455
 FT DOMAIN 28 250
 FT TRANSMEM 251 272
 FT TRANSMEM 278 299
 FT TRANSMEM 312 333
 FT DOMAIN 334 420
 FT TRANSMEM 421 442
 FT CARBOHYD 37 37
 FT CARBOHYD 137 137
 FT DISULFID 165 179
 FT MOTAGEN 91 91
 FT CONFLICT 435 435
 FT SEQUENCE 455 AA; 51753 MW; A270B43423B4086E CRC64;

Query Match 83.3%; Score 5; DB 1; Length 455;
 Best local Similarity 100.0%; Pred. No. 28;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 REPAIR 5
 DB 199 REPAIR 203

RESULT 10
 ID GAAL_BOVIN STANDARD; PRT; 456 AA.
 AC P08219;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-1 SUBUNIT PRECURSOR (GABA(A)
 DE RECEPTOR).
 GN GABRA1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Brain;
 RX MEDLINE=87258251; PubMed=3037384;
 RA Schofield P.R., Darlison M.G., Fujita N., Fujita N.,
 Rodriguez H., Rhee L.M., Ramchandran J., Reale V., Glencorse T.A.,
 Seeburg P.H., Barnard E.A.;
 RT "Sequence and functional expression of the GABA A receptor shows a
 RT ligand-gated receptor super-family.";
 RL Nature 328:221-227(1987).
 CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE
 CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE
 CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE
 CC CHANNEL.
 CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)
 CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: CEREBELLAR GRANULE CELLS, PURKINJE CELLS AND
 CC STRIATE/BASKET CELLS.
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05177; CAA29189.1; -.
DR PIR; A27142; A27142.
DR InterPro; IPR001175; Neur_channel.
DR Pfam; PF00065; neur_chan; 2.
DR PRINTS; PRO0252; NRIONCHANNEL.
DR PRINTS; PRO0253; GABARECEPTR.
DR PRINTS; PRO1079; GABARALPHA.
DR PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; signal;
KW Transmembrane; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 456
FT DOMAIN 28 251
FT TRANSMEM 252 273
FT TRANSMEM 279 300
FT TRANSMEM 313 334
FT DOMAIN 335 421
FT TRANSMEM 422 443
FT CARBOHYD 38 38
FT CARBOHYD 138 138
FT DISULFID 166 180
FT SEQUENCE 456 AA; 51655 MW; AFD4AA144B382CC2 CRC64;

```

```

Query Match 83.3%; Score 5; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 REPAR 5
Db 200 REPAR 204

RESULT 11
GAL_HUMAN STANDARD: PRT; 456 AA.
ID GAL_HUMAN
AC P14867;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-1 SUBUNIT PRECURSOR (GABA(A)
DE RECEPTOR).
DE GABRA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89153582; PubMed=2465923;
RA Seeburg P.H.; Pritchett D.B.; Sontheimer H.; Kettenmann H.;
RT "Sequence and expression of human GABA receptor alpha 1 and beta 1
RT subunits.";
RL FEBS Lett. 244:361-364(1989).
RN [2]
RP SEQUENCE OF 1-364 FROM N.A.
RC TISSUE=Cerebellum;
RX MEDLINE=89050079; PubMed=2847710;
RA Garrett K.M.; Duman R.S.; Salto N.; Blume A.J.; Vittek M.P.;
RA Tallman J.F.;
RT "Isolation of a cDNA clone for the alpha subunit of the human GABA-A
RT receptor.";

```

```

RL Biochem. Biophys. Res. Commun. 156:1039-1045(1988).
CC -! FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE
CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE
CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE
CC CHANNEL.
CC -! SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)
CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -! SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13584; CAA31925.1; -.
DR EMBL; X14766; CAA32874.1; -.
DR PIR; A31588; A31588.
DR PIR; S03332; S03332.
DR MIM; 137160; -.
DR InterPro; IPR001175; Neur_channel.
DR Pfam; PF00065; neur_chan; 2.
DR PRINTS; PRO0252; NRIONCHANNEL.
DR PRINTS; PRO0253; GABARECEPTR.
DR PRINTS; PRO1079; GABARALPHA.
DR PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; signal;
KW Transmembrane; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 456
FT DOMAIN 28 251
FT TRANSMEM 252 273
FT TRANSMEM 279 300
FT TRANSMEM 313 334
FT DOMAIN 335 421
FT TRANSMEM 422 443
FT CARBOHYD 38 38
FT CARBOHYD 138 138
FT DISULFID 166 180
FT CONFLICT 122 122
FT CONFLICT 128 140
FT CONFLICT 204 204
FT CONFLICT 315 315
FT CONFLICT 362 364
FT SEQUENCE 456 AA; 51771 MW; OCF0A780DF20D4E CRC64;

```

```

Query Match 83.3%; Score 5; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 REPAR 5
Db 200 REPAR 204

RESULT 12
SELA_PSEAE
ID SELA_PSEAE STANDARD: PRT; 468 AA.
AC Q9HV01;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE L-SERYL-TRNA(SEC) SELENIUM TRANSFERASE (EC 2.9.1.1) (CYSTEINYLT-
DE TRNA(SEC) SELENIUM TRANSFERASE) (SELENCYSTEINE SYNTHASE)
DE (SELENCYSTEINYLT-TRNA(SEC) SYNTHASE).
GN SELA OR PA4808.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

```

```

OC Pseudomonas.
CC NCBI_TaxID=287;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-ATCC 15692 / PAOI;
CC MEDLINE=2043737; PubMed=10984043;
CC Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
CC Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
CC Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
CC Brody L.L., Coulter S.N., Folger K.R., Kas A., Labdig K., Lim R.M.,
CC Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
CC Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
CC "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
CC opportunistic pathogen.";
CC Nature 406:959-964(2000).
CC -1- FUNCTION: CONVERTS SERYL-TRNA(SEC UCA) TO SELENOCYSTEINYL-TRNA(SEC
CC UCA) DURING SELENOPROTEIN BIOSYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-SERYL-TRNA(SEC) + SELENOPHOSPHATE = L-
CC SELENOCYSTEINYL-TRNA(SEC) + H(2)O + ORTHOPHOSPHATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SELA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE004893; AAG08194.1;
CC KW Transferase; Pyridoxal phosphate; Selenium; Complete proteome.
CC FT BINDING 291 291 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SQ SEQUENCE 468 AA; 50291 MW; 3ED90F21ED59723F CRC64;

Query Match 83.3%; Score 5; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
DB 42 REPAR 46

RESULT 13
PROL LEIEN
ID PROL LEIEN STANDARD; PRT; 567 AA.
AC P13665;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE PROBABLE TRANSPORT PROTEIN (LTP).
GN PRO-1.
OS Leishmania enriettii.
CC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
CC NCBI_TaxID=5663;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=90046659; PubMed=2813352;
CC Cairns B.R., Collard M.W., Landtgar S.M.;
CC "Developmentally regulated gene from Leishmania encodes a putative
CC membrane transport protein.";
CC Proc. Natl. Acad. Sci. U.S.A. 86:7682-7686(1989).
CC -1- FUNCTION: PROBABLE MEMBRANE TRANSPORT PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: PREDOMINANTLY FOUND IN PROMASTIGOTES.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: M26229; AAA29256.1;
CC DR PIR: A33974; A33974.
CC DR InterPro: IPR003663; Sugar_transport.
CC DR InterPro: IPR003662; sub_transport.
CC DR Pfam: PF00083; sugar_tr_1.
CC DR PRINTS: PR00171; SUGARTRANSPORT.
CC DR PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Transport; Sugar transport; Transmembrane.
CC SQ SEQUENCE 567 AA; 61451 MW; D6F8BA82AA4B62B CRC64;

Query Match 83.3%; Score 5; DB 1; Length 567;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
DB 19 EPARD 23

RESULT 14
FLIF_RHOSH
ID FLIF_RHOSH STANDARD; PRT; 570 AA.
AC Q53151;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FLAGELLAR M-RING PROTEIN.
GN FLIF.
OS Bacteroides spheroideis (Rhodopseudomonas spheroideis).
CC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
CC Rhodobacter.
CC NCBI_TaxID=1063;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=WS8;
CC Goodfellow I.G., Woolley K.J., Sockett R.E.S.;
CC Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE M RING MAY BE ACTIVELY INVOLVED IN ENERGY
CC TRANSDUCTION (BY SIMILARITY).
CC -1- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE
CC FLAGELLAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)
CC MOUNTED ON A CENTRAL ROD. THE M RING IS INTEGRAL TO THE INNER
CC MEMBRANE OF THE CELL AND MAY BE CONNECTED TO THE FLAGELLAR ROD
CC VIA THE S RING. THE S (SUPRAMEMBRANE RING) LIES JUST DISTAL TO
CC THE M RING. THE L AND P RINGS LIE IN THE OUTER MEMBRANE AND THE
CC PERIPLASMIC SPACE, RESPECTIVELY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, MEMBRANE-ASSOCIATED (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FLIF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X38692; CAA67251.1;
CC DR InterPro: IPR000067; Flgtrng_FLIF.
CC DR InterPro: IPR002920; YscJ_FLIF.
CC DR Pfam: PF01514; YscJ_FLIF; 1.
CC DR PRINTS: PR01009; FLGMRINGFLIF.
CC KW Flagella; Membrane.
CC SQ SEQUENCE 570 AA; 60675 MW; 87A6BF11C3DA6A46 CRC64;

```

Query Match 83.3%; Score 5; DB 1; Length 570;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: February 26, 2002, 01:53:55
Job time: 424 sec

OY 1 REPAR 5
|||||
DB 61 REPAR 65

RESULT 15
REF2P_DROME STANDARD: PRT; 599 AA.
AC P14199;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE REF(2)P PROTEIN.
GN REF(2)P.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=9005939; PubMed=2510997;
RA Dezelic S., Bras F., Contamine D., Lopez-Ferber M., Segretain D.,
RA Teninges D.;
RT "Molecular analysis of ref(2)P, a Drosophila gene implicated in sigma
rhodovirus multiplication and necessary for male fertility.";
RL EMO J. 8:3437-3446(1989).
CC -1- FUNCTION: IMPLICATED IN SIGMA RHADOVIRUS MULTIPLICATION AND
NECESSARY FOR MALE FERTILITY.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X16993; CA34861.1; ALT_SEQ.
DR PIR; S06785; S06785.
DR FlyBase: FBgn0003231; ref(2)P.
DR InterPro: IPR000270; OPR.
DR InterPro: IPR000449; UBA.
DR InterPro: IPR000433; Znf_ZZ.
DR Pfam: PF00564; OPR; 1.
DR Pfam: PF00627; UBA; 1.
DR Pfam: PF00569; ZZ; 1.
DR SMART: SM0016; OPR; 1.
DR SMART: SM00165; OPR; 1.
DR SMART: SM00291; Znf_ZZ; 1.
KW Zinc-finger; DNA-binding; Metal-binding; Nuclear protein; Repeat.
FT DOMAIN 386 413 3 X 8 AA REPEATS OF S-A-N-Q-S-X-X-P.
FT REPEAT 386 393 1.
FT REPEAT 399 406 2.
FT REPEAT 407 413 3.
FT REPEAT 407 413 3.
FT ZN_FING 140 153 C2H2-TYPE (ATYPICAL).
SO SEQUENCE 599 AA; 65735 MW; FE3C1E6C3FBB20FD CRC64;

Query Match 83.3%; Score 5; DB 1; Length 599;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
|||||
DB 211 EPARD 215

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:50:47 ; Search time 79.46 Seconds
(without alignments)
5.593 Million cell updates/sec

Title: US-09-819-136-2_COPY_283_288

Perfect score: 6
Sequence: 1 REPARD 6

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.GeneSeq_1101.*
1: /SID2/gcgdata/geneSeq/geneSeq/AA1980.DAT:*
2: /SID2/gcgdata/geneSeq/geneSeq/AA1981.DAT:*
3: /SID2/gcgdata/geneSeq/geneSeq/AA1982.DAT:*
4: /SID2/gcgdata/geneSeq/geneSeq/AA1983.DAT:*
5: /SID2/gcgdata/geneSeq/geneSeq/AA1984.DAT:*
6: /SID2/gcgdata/geneSeq/geneSeq/AA1985.DAT:*
7: /SID2/gcgdata/geneSeq/geneSeq/AA1986.DAT:*
8: /SID2/gcgdata/geneSeq/geneSeq/AA1987.DAT:*
9: /SID2/gcgdata/geneSeq/geneSeq/AA1988.DAT:*
10: /SID2/gcgdata/geneSeq/geneSeq/AA1989.DAT:*
11: /SID2/gcgdata/geneSeq/geneSeq/AA1990.DAT:*
12: /SID2/gcgdata/geneSeq/geneSeq/AA1991.DAT:*
13: /SID2/gcgdata/geneSeq/geneSeq/AA1992.DAT:*
14: /SID2/gcgdata/geneSeq/geneSeq/AA1993.DAT:*
15: /SID2/gcgdata/geneSeq/geneSeq/AA1994.DAT:*
16: /SID2/gcgdata/geneSeq/geneSeq/AA1995.DAT:*
17: /SID2/gcgdata/geneSeq/geneSeq/AA1996.DAT:*
18: /SID2/gcgdata/geneSeq/geneSeq/AA1997.DAT:*
19: /SID2/gcgdata/geneSeq/geneSeq/AA1998.DAT:*
20: /SID2/gcgdata/geneSeq/geneSeq/AA1999.DAT:*
21: /SID2/gcgdata/geneSeq/geneSeq/AA2000.DAT:*
22: /SID2/gcgdata/geneSeq/geneSeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5	83.3	107	21	Human ORFX ORF2288
2	5	83.3	136	22	Pseudomonas stutzeri
3	5	83.3	142	22	Human zaipha31. H
4	5	83.3	144	21	Human pancreatic c
5	5	83.3	154	19	S. pneumoniae putr
6	5	83.3	172	17	Human deoxycytidin
7	5	83.3	211	20	Human breast tumor
8	5	83.3	245	16	Hepatitis GB virus
9	5	83.3	245	16	Hepatitis GB virus
10	5	83.3	261	19	Human disease rela
11	5	83.3	277	21	Arbidopists thalia

12	5	83.3	291	22	AAW38843
13	5	83.3	291	22	AAW3694
14	5	83.3	295	21	AAW51904
15	5	83.3	299	21	AAW51903
16	5	83.3	302	19	AAW59495
17	5	83.3	303	21	AAW09218
18	5	83.3	303	22	AAW40629
19	5	83.3	312	21	AAW09217
20	5	83.3	325	21	AAW51902
21	5	83.3	329	21	AAW51901
22	5	83.3	389	19	AAW52823
23	5	83.3	410	19	AAW77354
24	5	83.3	419	21	AAW36455
25	5	83.3	435	9	AAW81258
26	5	83.3	456	19	AAW77355
27	5	83.3	456	22	AAW39294
28	5	83.3	462	22	AAW41080
29	5	83.3	500	19	AAW77356
30	5	83.3	500	20	AAW02185
31	5	83.3	500	22	AAW86167
32	5	83.3	510	22	AAW04829
33	5	83.3	516	21	AAW49947
34	5	83.3	529	22	AAW95631
35	5	83.3	552	19	AAW48873
36	5	83.3	568	16	AAW5496
37	5	83.3	573	19	AAW48874
38	5	83.3	618	19	AAW52827
39	5	83.3	622	17	AAW92702
40	5	83.3	756	21	AAW70975
41	5	83.3	867	21	AAW15739
42	5	83.3	876	20	AAW41293
43	5	83.3	876	22	AAW36597
44	5	83.3	906	21	AAW15740
45	5	83.3	1466	20	AAW41091

ALIGNMENTS

RESULT 1	AAW42524	Human ORFX ORF2288 polypeptide sequence SEQ ID NO:4576.
ID	AAW42524	standard; Protein: 107 AA.
AC	AAW42524:	
DT	08-FEB-2001	(first entry)
DE	Human ORFX ORF2288 polypeptide sequence SEQ ID NO:4576.	
KW	Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;	Human polypeptide
KW	vulnerary; antiproliferative; antiparasitism; noctropic; neuroprotective;	Human oxidoreducta
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	Murine 3GNT protei
KW	hypotensive; dermatological; immunosuppressive; antidiabetic;	Human 3GNT protei
KW	antiviral; antibacterial; antifungal; antipneumatic; antihypoid;	Human 3GNT protei
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;	Human 3GNT protei
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	Human 3GNT protei
KW	cardiovascular disease; diabetes mellitus; hypochylodism; SCID; AIDS;	Human 3GNT protei
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	Human 3GNT protei
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	Human 3GNT protei
KW	allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;	Human 3GNT protei
KW	bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;	Human 3GNT protei
KW	thrombosis; contraceptive.	Human 3GNT protei
KW	Homo sapiens.	Human 3GNT protei
OS	MO20005473-A2.	Human 3GNT protei
PN	05-OCT-2000.	Human 3GNT protei
PD	31-MAR-2000; 2000MO-US08621.	Human 3GNT protei
PF	31-MAR-1999; 99US-0127607.	Human 3GNT protei
XX		Human 3GNT protei
PR		Human 3GNT protei

```

PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI: 2000-602362/57.
XX N-PSDB; AAC76733.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 3773; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotrophic; vulnary;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
CC osteoplastic; anticoagulant; antithrombotic; immunosuppressive;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antitumor; antibacterial; antiviral; antifungal; antineoplastic;
CC antihypertensive; antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 107 AA;

Query Match 83.3%; Score 5; DB 21; Length 107;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAIR 5
   |||||
Db 31 repair 35

RESULT 2
AAE05855
ID AAE05855 standard; Protein; 136 AA.
XX
AC AAE05855;
XX
DT 24-SEP-2001 (first entry)
XX
DE Pseudomonas stutzeri open reading frame-G (ORF-G) protein.
XX
KW Open reading frame-G; ORF-G; pyridine-2,6-bis (thiocarboxylate);
KW PDRC; environmental remediation; phytoextraction; bioaccumulation;
KW water purification; solution mining mobilisation; immobilisation;
KW detoxification; redox state modifier; metal ion reactivity;
KW carbon tetrachloride; metal.
XX
XX Pseudomonas stutzeri.
OS
XX WO200153309-A1.
XX
PD 26-JUL-2001.
XX

```

```

PF 19-JAN-2001; 2001WO-US02386.
XX
XX 20-JAN-2000; 2000US-0177251.
XX
PA (IDAH-) IDAHO RES FOUND INC.
XX (PASZ/) PASZCZYNSKI A.
XX (SEBA/) SEBAT J L.
XX
PI Paszczyński A, Sebat JL, Lewis TA, Crawford RL, Cortese MS;
XX
DR WPI: 2001-465361/50.
XX N-PSDB; AAD11175.
XX
PT New isolated nucleic acids, useful for producing enzymes required to
PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for
PT reducing the amount of metal or carbon tetrachloride in a substrate,
PT e.g. soil or water -
XX
PS Example 1; Page 108; 172pp; English.
XX
CC The present sequence is a Pseudomonas stutzeri open reading frame-G
CC (ORF-G) protein. The Pseudomonas stutzeri genome includes ORFs that
CC encode enzymes required for biosynthesis of pyridine-2,6-bis
CC (thiocarboxylate) (PDRC). The ORFs encoding PDRC are especially useful
CC in environmental remediation methods, e.g. phytoextraction,
CC bioaccumulation, water purification, waste water purification, solution
CC mining mobilisation, immobilisation, detoxification, redox state
CC modifier or modification of metal ion reactivity. In particular, the
CC ORFs are useful for degrading carbon tetrachloride and removing metals
CC from substrates, e.g. soil or water.
XX
SQ Sequence 136 AA;

Query Match 83.3%; Score 5; DB 22; Length 136;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAIR 5
   |||||
Db 114 repair 118

RESULT 3
AAB50867
ID AAB50867 standard; Protein; 142 AA.
XX
AC AAB50867;
XX
DT 19-MAR-2001 (first entry)
XX
DE Human zalpha31.
XX
KW Human; zalpha31; alpha helical protein-31; expressed sequence tag; EST;
KW cytosolic; antiinflammatory; antithrombotic; antineoplastic;
KW immunosuppressive; antibacterial; osteoplastic;
KW antiatherosclerotic; antiproliferative; dermatological; anti-HIV;
KW gene therapy; osteoporosis; Paget's disease; hyperparathyroidism;
KW acute pancreatitis; gastrointestinal disorder; inflammation;
KW heart disorder; viral infection.
XX
XX Homo sapiens.
OS
XX WO200073458-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14795.
XX
PR 28-MAY-1999; 99US-0136485.
XX
XX (ZYMO) ZYMOGENETICS INC.
XX

```

PI Conklin DC:
XX
XX WPI: 2001-070967/08.
DR N-PSDB: AAC91386, AAC91387.
XX
XX Novel 4-helix bundle cytokine, Zalpha31, useful for regulating the
PT function of immune system and for treating thyroid, adrenal, lymphoid,
PT inflammatory, pancreatic, blood or bone disorders
XX
XX Claim 10; Page 99-100; 11pp; English.
PS
XX
XX The present sequence is given in a specification relating to
CC polynucleotide and polypeptide molecules for mammalian secreted alpha
CC helical protein-31 (Zalpha31). The polypeptides are a novel four-helix
CC bundle cytokine and may be used to regulate the functioning of the
CC immune system. The polypeptides are used to identify and isolate
CC receptors involved in spermatogenesis, steroidogenesis, testicular
CC differentiation and regulatory control of the hypothalamic-pituitary
CC gonadal axis, thyroid, heart and adrenal function. They are useful for
CC treating disorders of the reproductive system, thyroid, adrenal, heart
CC and immunological systems. Zalpha31 polypeptides, or antagonists are
CC useful in the treatment of e.g. osteoporosis, Paget's disease, and
CC hyperparathyroidism, acute pancreatitis and gastrointestinal disorders
CC and as analgesics, especially for bone pain. They can be used to treat
CC atherosclerosis, pelvic inflammatory disease, (PID), psoriasis, The
CC arthritis, eczema, scleroderma and other inflammatory diseases. The
CC polypeptides, nucleic acids and/or antibodies can be used to treat heart
CC disorders. They can limit infarct size following a heart attack, aid in
CC recovery after heart transplantation, treat complications related to
CC poor circulation e.g. diabetic foot ulcers, treat stroke, improve
CC cardiac function, induce skeletal muscle neogenesis and/or hyperplasia,
CC be used in kidney regeneration and/or for treating systemic and
CC pulmonary hypertension. Zalpha31 can be used for treating viral
CC leukaemias, acquired immunodeficiency syndrome (AIDS) or gastrointestinal
CC viral infections.
XX
XX Sequence 142 AA:
SQ

Query Match 83.3%; Score 5; DB 22; Length 142;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
| | | | |
Db 15 epard 19

RESULT 4
AAB54229 standard; Protein; 144 AA.
ID AAB54229;
XX
XX AAB54229;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:661.
XX
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytosolic; neuroprotective;
KW neurotropic; immunomodulatory; relaxant; contraceptive; gynecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neutral; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX
XX Homo sapiens.
OS
XX
XX WO200055320-A1.
PN
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05989.
PF

XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX WPI: 2000-579444/54.
DR
XX N-PSDB: AAC98994.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 11; Page 1117-1118; 1379pp; English.
PS
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytosolic,
CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,
CC gynecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
XX Sequence 144 AA:
SQ

Query Match 83.3%; Score 5; DB 21; Length 144;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
| | | | |
Db 17 epard 21

RESULT 5
AAB38636 standard; Protein; 154 AA.
ID AAB38636;
XX
XX AAB38636;
XX
DT 09-NOV-1998 (first entry)
XX
DE S. pneumoniae putrescine transport ATP-binding protein.
XX
XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis.
XX
XX Streptococcus pneumoniae.
OS
XX
XX WO9743303-A1.
PN
XX 20-NOV-1997.
PD
XX

PF 14-MAY-1997; 97WO-US07950.
 XX
 PR 14-MAY-1996; 96US-0017670.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1998-008793/01.
 DR N-PSDB: AAT98682.
 XX
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 XX
 PS Claim 12; Pages 393-394; 483pp; English.
 XX
 CC This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with an E. coli protein, is a putrescine transport ATP-
 CC binding protein, and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 CC
 XX
 SO Sequence 154 AA;
 XX
 QY 1 REPAIR 5
 DB 136 repair 140
 XX
 RESULT 6
 ID AAW00482
 XX AAW00482 standard; Protein; 172 AA.
 AC AAW00482;
 XX
 DT 27-SEP-1996 (first entry)
 XX
 DE Human deoxycytidine kinase 2.
 XX
 KW Deoxycytidine kinase 2; hdk2; malignancy; cancer; virus infection;
 KW nucleoside analogue; drug activation; therapy; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PI
 PI Key Location/Qualifiers
 FH Domain 28..34
 FT /label= ATP-binding_domain
 XX
 PN WO9621724-A1.
 XX
 PD 18-JUL-1996.

XX
 PF 13-JAN-1995; 95WO-US00532.
 XX
 PR 13-JAN-1995; 95WO-US00532.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Kirkness EF, Wei Y;
 XX
 DR WPI: 1996-342279/34.
 DR N-PSDB: AAT31065.
 XX
 PT DNA encoding human deoxycytidine kinase 2 - useful in treatment of
 PT malignancies and viral infections
 XX
 PS Claim 1; Page 37; 56pp; English.
 XX
 CC Human mature deoxycytidine kinase 2 (hdk2) (AAW00482) plays a
 CC physiological role in the maintenance of normal deoxyribonucleotide
 CC pools, and a key role in the phosphorylation of antineoplastic and
 CC antiviral nucleoside analogues. It was identified by homology to
 CC mouse dck1. It can be expressed in e.g. E. coli, COS or Sf9 cells
 CC utilising vectors incorporating hdk2 encoding sequences (see also
 CC AAT31065). The polypeptide is useful for treating malignancies and
 CC viral infections, to screen for antagonists (useful in the
 CC treatment of immunodeficiency diseases), and to raise antibodies.
 CC
 XX
 SO Sequence 172 AA;
 XX
 QY 1 REPAIR 5
 DB 84 repair 88
 XX
 RESULT 7
 ID AAY48508
 XX AAY48508 standard; Protein; 211 AA.
 AC AAY48508;
 XX
 DT 08-DEC-1999 (first entry)
 XX
 DE Human breast tumour-associated protein 53.
 XX
 KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 KW medicaments; gene therapy; treatment; fat metabolism.
 XX
 OS Homo sapiens.
 XX
 PI DE19813835-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 20-MAR-1998; 98DE-1013835.
 XX
 PR 20-MAR-1998; 98DE-1013835.
 XX
 PA (META-) METAGEN GES GENOFORSCHUNG MBH.
 XX
 PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 PI WPI: 1999-528979/45.
 DR N-PSDB: AAZ33589.
 XX
 PT Human nucleic acid sequences and protein products from normal breast
 PT tissue, useful for breast cancer therapy
 XX
 PS Claim 28; 182; 206pp; German.

XX This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytostatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for treatments for gene
CC therapy to treat breast cancer and for treating illnesses associated
CC with fat metabolism. AA148456-Y48539 represent protein fragments encoded
CC by the expressed sequence tags described in the method of the invention.
XX

SQ Sequence 211 AA;

Query Match 83.3%; Score 5; DB 20; Length 211;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAIR 5
|||||
DB 15 repair 19

RESULT 8
AAR81413
ID AAR81413 standard; Protein: 245 AA.
XX
AC AAR81413;
XX
DT 02-JUL-1996 (first entry)
XX
DE Hepatitis GB virus (HGBV) clone 2 protein prod.
XX
KW Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW reagents; non-A; non-B; non-C; non-D; non-E; clone 2;
KW tamarin; infected plasma; lambda phage; CDNA library.
XX
OS Hepatitis GB virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 59 /note= "corresponding codon STOP codon"
FT Misc-difference 91 /note= "corresponding codon STOP codon"
XX
PN W09521922-A2.
XX
PD 17-AUG-1995.
XX
PE 14-FEB-1995; 95WO-US02118.
XX
PR 27-JAN-1995; 95US-0344557.
PR 14-FEB-1994; 94US-0196030.
PR 13-MAY-1994; 94US-0242654.
PR 29-JUL-1994; 94US-0283314.
PR 23-NOV-1994; 94US-0344185.
PR 23-NOV-1994; 94US-0344190.
XX
PA (ABBO) ABBOTT LAB.
PI Buljk SL, Dawson GJ, Desai SM, Erker JC, Leary TP,
PI Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
PI Simons JN;
XX
XX WPI: 1995-293123/38.
DR N-PSDB: AAT00042.
XX
XX Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT for diagnosis and therapy of hepatitis GB virus
XX
XX Example 5; Pages 199-200; 661pp; English.
XX
XX Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC

CC infected tamarin plasma, using standard procedures, was used to
CC prepare a lambda phage HGBV cDNA library. The cDNA clone AA100042,
CC which encodes the proteins AAR81411-16 (the 6 possible reading
CC frames), was rescued from the lambda phage, searched against a
CC sequence database and found to be an unique HGBV sequence.
CC Reagents which comprise the HGBV DNA, or its protein prods. can
CC be used for the diagnosis, therapy or in a vaccine to prevent
CC HGBV infection.
XX

SQ Sequence 245 AA;

Query Match 83.3%; Score 5; DB 16; Length 245;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAIR 5
|||||
DB 76 repair 80

RESULT 9
AAB08996
ID AAB08996 standard; Protein: 245 AA.
XX
AC AAB08996;
XX
DT 30-AUG-2000 (first entry)
XX
DE Hepatitis GB virus protein sequence SEQ ID NO:39.
XX
KW Hepatitis GB virus; HGBV; diagnosis; therapeutic; immunogenic;
KW infection; detection; characterisation; hepatitis.
XX
OS Hepatitis GB virus.
XX
PN US6051374-A.
XX
PD 18-APR-2000.
XX
PE 07-JUN-1995; 95US-0488445.
XX
PR 14-FEB-1994; 94US-0196030.
PR 13-MAY-1994; 94US-0242654.
PR 29-JUL-1994; 94US-0283314.
PR 23-NOV-1994; 94US-0344185.
PR 23-NOV-1994; 94US-0344190.
PR 30-JAN-1995; 95US-0377557.
XX
PA (ABBO) ABBOTT LAB.
PI Dawson GJ, Leary TP, Muerhoff AS, Pilot-Matias TJ, Buljk SL,
PI Mushahwar IK, Simons JN, Desai SM, Erker JC, Schlauder GG;
PI Simons JN;
XX
XX WPI: 2000-338307/29.
DR
XX
PT Detecting target hepatitis GB virus nucleic acid in a test sample
PT suspected of containing HGBV comprises reacting the test sample the
PT HGBV polynucleotide probe and detecting the complex that contains
PT target HGBV -
XX
PS Example 5; Column 159-162; 369pp; English.
XX
XX The present invention describe a method for detecting target
CC hepatitis GB virus (HGBV) nucleic acid (THN) in a test sample (T) with a
CC suspected of containing HGBV. The method involves reacting (T) with a
CC HGBV polynucleotide probe (II) containing 15 contiguous nucleotides, and
CC which selectively hybridises to the HGBV genome or its full complement,
CC and detecting the complex that contains THN, indicating the presence of
CC target HGBV. The method is used for detecting target HGBV nucleic acid
CC in the test sample suspected of containing HGBV and for characterisation
CC of newly ascertained etiological agent of non-A, non-B, non-C, non-D and
CC non-E hepatitis causing agents collectively termed as hepatitis GB
CC

CC virus. AAA55270 to AAA55489 and AAB08965 to AAB09480 represent nucleotide
CC and protein sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 245 AA;

Query Match 83.3%; Score 5; DB 21; Length 245;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAR 5
Db 76 repair 80

RESULT 10
AAW70493
ID AAW70493 standard; Protein; 261 AA.
AC AAW70493;
XX
DT 29-DEC-1998 (first entry)
XX
DE Human disease related nucleotide kinase-1 (DRNK-1) protein sequence.
XX
KW Human disease related nucleotide kinase-1; DRNK-1; deoxyguanosine kinase;
KW p21ras; cell proliferation; oncogenesis; cancer; gene therapy;
KW immune disorder; neurological dysfunction.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT /note= "Mitochondrial localisation signal"
FT Modified-site 8
FT /note= "N-myristoylated"
FT Domain 28..34
FT /note= "ATP/GTP binding site (p-loop)"
FT Modified-site 141
FT /note= "N-glycosylated"
XX
PN US5817482-A.
XX
PD 06-OCT-1998.
XX
PF 20-JUN-1997; 97US-0879561.
XX
PR 20-JUN-1997; 97US-0879561.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Corley NC, Guegler KJ, Hawkins PR, Hillman JL;
XX
DR WPI: 1998-556388/47.
DR N-PSDB; AAV33481.
XX
PT Nucleic acids encoding deoxyguanosine kinase - useful for
PT recombinant production of the enzyme for treating diseases caused by
PT lack of the enzyme e.g. cancers caused through loss of enzyme
PT function
XX
PS Claim 1; Fig 1A-1C; 53pp; English.
XX
XX The present sequence represents a human disease related nucleotide
CC kinase-1 (DRNK-1) protein sequence. The DRNK-1 encoding DNA sequence
CC was first identified in Incyte Clone 56821 from the fibroblast cDNA
CC library (FIBRNOT 01). The DRNK-1 DNA sequence is useful for the
CC production of the corresponding recombinant enzyme. The invention
CC provides DNK enzymes, which are deoxyguanosine kinases, which catalyse
CC the transfer of a terminal phosphate from adenosine triphosphate (ATP)
CC or guanine triphosphate (GTP) to guanosine or guanine in the regulation
CC of cellular levels of GTP and its corresponding nucleoside triphosphate.

CC As GTP levels are known to control the activity of certain oncogenic
CC proteins e.g. p21ras, a protein involved in cell proliferation and
CC oncogenesis, suppression of the enzyme activity causes high ratios of
CC GTP:GDP, promoting oncogenesis. Therefore, diseases (e.g. cancers,
CC immune disorders and neurological dysfunction) related to this lack of
CC activity may be prevented or treated with the recombinant enzyme, or by
CC gene therapy based strategies. Anti-sense constructs of the DRNK
CC encoding nucleic acids may also be used for inhibition of over-expression
CC of the enzyme.
XX
SQ Sequence 261 AA;

Query Match 83.3%; Score 5; DB 19; Length 261;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAR 5
Db 84 repair 88

RESULT 11
AAG28283
ID AAG28283 standard; Protein; 277 AA.
XX
AC AAG28283;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 33443.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0126834.
PR 08-APR-1999; 99US-0126714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0148722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154179.
PR 22-SEP-1999; 99US-0155137.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match

83.3%; Score 5; DB 21; Length 277;
 Best Local Similarity 100.0%; Pred. No. 65;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
 |||||
 Db 22 repar 26

RESULT 12

AAM3843
 ID AAM3843 standard; Protein; 291 AA.

AC AAM3843;

DE 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 1988.

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukemia.

XX Homo sapiens.

XX WO20015312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR MPI: 2001-442253/47.

DR N-Psdb; AAI57999.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PS Such as central nervous system injuries -

XX Example 3; SEQ ID NO 1988; 10078pp; English.

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM3842-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression.

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: the sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 291 AA;

Query Match 83.3%; Score 5; DB 22; Length 291;
 Best Local Similarity 100.0%; Pred. No. 67;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EPARD 6
 |||||
 Db 219 epard 223

RESULT 13

AAB73694
 ID AAB73694 standard; Protein; 291 AA.

AC AAB73694;

DE 11-SEP-2001 (first entry)

XX Human oxidoreductase protein ORP-27.

KW Human oxidoreductase protein; ORP; cell proliferative disorder;
 KW arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;
 KW diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhea;
 KW osteoporosis; metabolic disorder; obesity; phenylketonuria;
 KW hypercholesterolaemia; reproductive disorder; infertility;
 KW ovulatory defect; menstrual cycle defect; endometriosis;
 KW polycystic ovary disease; spermatogenesis disruption; impotence;
 KW neurological disorder; epilepsy; stroke; Alzheimer's disease;
 KW Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;
 KW meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;
 KW schizophrenic disorder; infection; autoimmune disorder;
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;
 KW allergy; Crohn's disease; atopic dermatitis; gout; multiple sclerosis;
 KW rheumatoid arthritis; ulcerative colitis; drug screening;
 KW toxicity screening; transgenic animal; SNP detection; gene therapy.

XX Homo sapiens.

XX WO20014448-A2.

XX 21-JUN-2001.

XX 07-DEC-2000; 2000WO-US33158.

XX 16-DEC-1999; 99US-0172367.

XX (INCY-) INCYTE GENOMICS INC.

PI Yue H, Lal P, Tang YT, Hillman JL, Baughn MR, Azimzai Y, Lu DM;

DR MPI: 2001-390245/41.

DR N-Psdb; AAB73694.

PT Novel human oxidoreductase protein (ORP) useful for diagnosing,

PT treating and preventing cell proliferative, neurological, viral,

PT reproductive and autoimmune/inflammatory disorders associated with

PS abnormal expression of ORP -

XX Claim 1; Page 120-121; 136pp; English.

CC Sequences AAB73668-AAB73694 represent 27 novel human oxidoreductase
 CC proteins, designated ORP-1 to ORP-27 respectively, and sequences
 CC AAB73668-AAB73694 represent cDNAs encoding ORP-1 to ORP-27. Human ORP
 CC proteins and nucleic acids are useful for diagnosing, treating or

CC preventing cell proliferative disorders (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, cancers); endocrine disorders (e.g., type I or II
CC diabetes mellitus, diabetes insipidus, dwarfism, hirsutism, amenorrhoea,
CC osteoporosis); metabolic disorders (e.g., obesity, phenylketonuria,
CC hypercholesterolemia); reproductive disorders (e.g., infertility,
CC ovulatory and menstrual cycle defects, endometriosis, polycystic
CC ovary disease, disruption of spermatogenesis, impotence); neurological
CC disorders (e.g., epilepsy, stroke, Alzheimer's disease, Huntington's
CC disease, Parkinson's disease, meningitis, Creutzfeldt-Jakob disease,
CC cerebellar palsy, muscular dystrophy, mood, anxiety and schizophrenic
CC disorders); viral, bacterial, fungal and parasitic infections; and
CC autoimmune/inflammatory disorders such as acquired immunodeficiency
CC syndrome (AIDS), allergies, asthma, Crohn's disease, atopic dermatitis,
CC gout, multiple sclerosis, rheumatoid arthritis or ulcerative colitis.
CC Human ORP proteins and nucleotides can be used to identify compounds
CC which modulate their activity or expression. ORP nucleic acid sequences
CC may also be used for assessing the toxicity of a test compound, to detect
CC upstream sequences such as promoters and regulatory elements, and to
CC create knock out or knock in animals or transgenic animals to model
CC human disease. Oligonucleotide primers derived from ORP gene sequences
CC may be used to detect single nucleotide polymorphisms (SNPs) and for
CC mapping the naturally occurring genomic sequences. Antibodies specific
CC for ORP proteins may be used in the diagnosis of disorders associated
CC with aberrant ORP expression, in assays to monitor patients being treated
CC with ORP or modulators thereof, and for assessing toxicity of potential
CC drugs.

XX Sequence 291 AA;

Query Match 83.3%; Score 5; DB 22; Length 291;

Best Local Similarity 100.0%; Pred. No. 67;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EPARD 6

|||||

Db 219 epard 223

RESULT 14

AAVS1904

ID AAVS1904 standard; Protein; 295 AA.

XX AAVS1904;

DT 15-JUN-2000 (first entry)

XX Murine 3GnT protein fragment.

DE UDP (uridine diphosphate)-N-acetylglucosaminyl; beta1,4 galactoside;

KW beta1,3-N-acetylglucosaminyl transferase; 3GnT; polylactosaminyl;

KM disaccharide production; oligosaccharide production; glycan;

KW glycoprotein; gene therapy; transgenic animal; murine.

XX Mus musculus.

OS WO200011190-A2.

PN 02-MAR-2000.

PF 06-AUG-1999; 99WO-CH00365.

PR 20-AUG-1998; 98CH-0001717.

XX (BERG/) BERGER E G.

PA (HENN/) HENNET T.

XX Berger EG, Hennet T;

PI WPI; 2000-224707/19.

DR New nucleic acid encoding a glucosaminyl transferase, used for in vivo

XX or in vitro synthesis of glycan chains containing a specific

PT or in vitro synthesis of glycan chains containing a specific

PT disaccharide repeat element -

XX Disclosure; Page 34; 34pp; German.

CC This invention describes a novel isolated nucleic acid (I) encoding
CC UDP (uridine diphosphate)-N-acetylglucosaminyl; beta1,4

CC galactoside; beta1,3-N-acetylglucosaminyl transferase (3GnT)

CC (polylactosaminyl type). 3GnTs are used, in vivo or in vitro, for

CC enzymatic production of di- or oligo-saccharides, particularly glycans,

CC containing the repeated disaccharide motif GlcNAc-beta1-3Gal. (I) is

CC useful for recombinant expression of 3GnTs, to alter adhesion and

CC recognition properties of transfected cells and to produce glycoproteins

CC substituted by polylactosaminoglycans, when introduced into cells that

CC express glycoproteins. Vectors containing (I) are used in gene therapy

CC and for production of transgenic animals. Fragments of (I) are used to

CC detect related sequences from other species and as antisense reagents.

CC This sequence represents a murine 3GnT protein fragment described in the

CC invention.

XX Sequence 295 AA;

Query Match 83.3%; Score 5; DB 21; Length 295;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REPAR 5

|||||

Db 139 repar 143

RESULT 15

AAVS1903

ID AAVS1903 standard; Protein; 299 AA.

XX AAVS1903;

DT 15-JUN-2000 (first entry)

XX Human 3GnT protein fragment.

DE UDP (uridine diphosphate)-N-acetylglucosaminyl; beta1,4 galactoside;

KW beta1,3-N-acetylglucosaminyl transferase; 3GnT; polylactosaminyl;

KM disaccharide production; oligosaccharide production; glycan;

KW glycoprotein; gene therapy; transgenic animal; human.

XX Homo sapiens.

OS WO200011190-A2.

PN 02-MAR-2000.

PF 06-AUG-1999; 99WO-CH00365.

PR 20-AUG-1998; 98CH-0001717.

XX (BERG/) BERGER E G.

PA (HENN/) HENNET T.

XX Berger EG, Hennet T;

PI WPI; 2000-224707/19.

DR New nucleic acid encoding a glucosaminyl transferase, used for in vivo

XX or in vitro synthesis of glycan chains containing a specific

PT disaccharide repeat element -

XX Disclosure; Page 33-34; 34pp; German.

CC This invention describes a novel isolated nucleic acid (I) encoding

CC UDP (uridine diphosphate)-N-acetylglucosaminyl; beta1,4

CC galactoside; beta1,3-N-acetylglucosaminyl transferase (3GnT)

CC (polylactosaminyl type). 3GnTs are used, in vivo or in vitro, for

CC enzymatic production of di- or oligo-saccharides, particularly glycans,
CC containing the repeated disaccharide motif GlcNAc β Gal. (1) is
CC useful for recombinant expression of 3GNTs, to alter adhesion and
CC recognition properties of transfected cells and to produce glycoproteins
CC substituted by polyactosaminoglycans, when introduced into cells that
CC express glycoproteins. Vectors containing (1) are used in gene therapy
CC and for production of transgenic animals. Fragments of (1) are used to
CC detect related sequences from other species and as antisense reagents.
CC This sequence represents a human 3Gnt protein fragment described in the
XX invention.
SO Sequence 299 AA;

Query Match 83.3%; Score 5; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 REPAR 5
|||
Db 143 repar 147

Search completed: February 26, 2002, 01:50:48
Job time: 1252 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:50:46 ; Search time 79.46 seconds
(without alignments)
5.593 Million cell updates/sec

Title: US-09-819-136-2_COPY_525_530

Perfect score: 6
Sequence: 1 EKRWK 6

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.GeneSeq_1101.*
1: /SID2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	163	21	AA12849
2	6	100.0	421	21	AA28824
3	5	83.3	13	22	AA97642
4	5	83.3	40	21	AA57567
5	5	83.3	52	18	AA08157
6	5	83.3	58	22	AA00689
7	5	83.3	70	18	AA18344
8	5	83.3	107	21	AA17299
9	5	83.3	111	21	AA37567
10	5	83.3	121	21	AA10185
11	5	83.3	132	21	AA01051

12	5	83.3	140	21	AA17298	Arabidopsis thalia
13	5	83.3	144	21	AA37566	Arabidopsis thalia
14	5	83.3	145	21	AA16726	Bacteriophage Dp-1
15	5	83.3	148	21	AA05046	Arabidopsis thalia
16	5	83.3	151	21	AA26405	Arabidopsis thalia
17	5	83.3	158	20	AA42130	Wheat Tryptophan s
18	5	83.3	166	21	AA15540	Arabidopsis thalia
19	5	83.3	166	21	AA20975	Arabidopsis thalia
20	5	83.3	166	21	AA52923	Arabidopsis thalia
21	5	83.3	167	21	AA15539	Arabidopsis thalia
22	5	83.3	167	21	AA20974	Arabidopsis thalia
23	5	83.3	167	21	AA52922	Arabidopsis thalia
24	5	83.3	172	21	AA26404	Arabidopsis thalia
25	5	83.3	183	21	AA20973	Arabidopsis thalia
26	5	83.3	184	21	AA05045	Arabidopsis thalia
27	5	83.3	186	21	AA52921	Arabidopsis thalia
28	5	83.3	187	21	AA10184	Arabidopsis thalia
29	5	83.3	187	21	AA15538	Arabidopsis thalia
30	5	83.3	194	21	AA17297	Arabidopsis thalia
31	5	83.3	198	21	AA37565	Arabidopsis thalia
32	5	83.3	202	21	AA35858	Arabidopsis thalia
33	5	83.3	205	21	AA35857	Arabidopsis thalia
34	5	83.3	218	21	AA35856	Arabidopsis thalia
35	5	83.3	225	21	AA48935	Arabidopsis thalia
36	5	83.3	230	21	AA29542	Arabidopsis thalia
37	5	83.3	235	21	AA49011	Arabidopsis thalia
38	5	83.3	251	15	AA53085	Transcription acti
39	5	83.3	258	22	AA04777	Human vesicle traf
40	5	83.3	305	21	AA29541	Arabidopsis thalia
41	5	83.3	306	21	AA48934	Arabidopsis thalia
42	5	83.3	306	21	AA15918	E. coli proliferat
43	5	83.3	306	22	AA69891	E. coli growth and
44	5	83.3	314	21	AA29540	Arabidopsis thalia
45	5	83.3	344	21	AA16071	Arabidopsis thalia

ALIGNMENTS

RESULT	1
AA12849	standard; Protein: 163 AA.
XX	
AC	AA12849:
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 12120.
XX	
KW	Protein identification: signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0133248.
PR 05-MAY-1999; 99US-0133248.
PR 06-MAY-1999; 99US-0133248.
PR 06-MAY-1999; 99US-0133248.
PR 06-MAY-1999; 99US-0133248.
PR 07-MAY-1999; 99US-0133248.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139415.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145146.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145918.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.


```
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          100.0%; Score 6; DB 21; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKRvKK 6
   |||||
Db 8 ekrvkk 13

RESULT 2
AAG28824 standard; Protein; 421 AA.
XX
XX AAG28824;
AC
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 34187.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130691.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
```

```
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134378.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140355.
PR 28-JUN-1999; 99US-0140623.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143524.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
```

PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	100.0%;	Score 6;	DB 21;	Length 421;
Best Local Similarity	100.0%;	Pred. No. 7.6;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1 EKRVRK 6
DB	8 ekrvrk 13

RESULT 3
AAB97642
ID AAB97642 standard; peptide; 13 AA.
XX AAB97642;
AC
XX
DT
XX
21-SEP-2001 (first entry)
DE Adenovirus core protein pV N-terminal nuclear localisation sequence.
XX
KW Adenovirus; core protein pV; nucleic acid packaging peptide gene therapy;
KW non-viral transfection; gene delivery; genetic vaccine;
KW in vitro transfection; neurodegenerative disease; cancer; tumour;
KW genetic disease; immune-related disorder; autoimmune disease;
KW immune hyperreactivity; allergy; nuclear localisation sequence; NLS;
KW N-terminal.
XX
OS Adenoviridae.
XX
PN WO200148233-A1.
XX
PD 05-JUL-2001.
XX
PF 12-DEC-2000; 2000WO-GB04767.
XX
PR 23-DEC-1999; 99GB-0030533.
XX
PA (MITS-) MITSUBISHI-TOKYO PHARM INC.
XX
PI Tagawa T, Miller AD, Perouzel E, Murray K, Manwell M, Alton E;
PI Matthews D, Russell W;
XX
DR WPI; 2001-441719/47.
XX
PT Nucleic acid delivery complex for delivering nucleic acids to cells
PT such as neuronal, cancer, epithelial cells; comprises cationic
PT lipid/protein/nucleic acid complex comprising viral packaging proteins
PT
PM
XX
PS
XX
Claim 10; Page 10; 71pp; English.
CC
CC The invention relates to a non-viral nucleic acid delivery vector
CC comprising a condensed polypeptide/nucleic acid complex and a cationic
CC lipid. The polypeptide/nucleic acid complex comprises a nucleic acid
CC sequence of interest and one or more viral nucleic acid packaging
CC polypeptides or derivatives thereof, capable of binding to and
CC condensing the nucleotide of interest, which is heterologous to the
CC polynucleotide to be transfected.

CC localisation sequence (NLS) and is thus able to deliver the nucleotide
CC of interest to the nucleus of a eukaryotic cell. Examples of viral
CC nucleic acid packaging polypeptides include viral core proteins such as
CC the adenoviral core proteins Mui, pV, and pVII. The non-viral nucleic
CC acid delivery vector and polypeptide/nucleic acid complex are useful in
CC variety of applications such as gene therapy. DNA vaccine delivery and
CC in vitro transfection studies. The delivery vectors/complexes are useful
CC for treating diseases of the peripheral or central nervous system such as
CC neurodegenerative diseases (e.g., motor neuron disease, familial
CC dysautonomia and infantile spinal muscular atrophy), late onset
CC neurodegenerative diseases such as Parkinson's and Alzheimer's diseases,
CC and damage to nervous tissue as a result of injury/trauma (including
CC stroke). They also are useful for treating solid tumours (e.g., cancers
CC of the breast, cervix, colon, lung, ovary, prostate gland, skin, stomach,
CC and bladder) and non-solid tumours (e.g., leukaemia and lymphoma);
CC diseases caused by inherited or somatic mutations, such as cystic
CC fibrosis and thalassemia; and immune-related disorders, such as organ
CC transplant rejection, autoimmune diseases (e.g., thyroiditis, insulinitis,
CC multiple sclerosis, hepatitis, Addison's disease, myasthenia gravis,
CC rheumatoid arthritis, and systemic lupus erythematosus) and immune
CC hyperreactivity (e.g., allergic reactions, in particular associated with
CC histamine production, and asthma). Sequences AAB97641 and AAB97642
CC represent, respectively, the C-terminal and N-terminal nuclear
CC localisation sequences (NLS) of the adenoviral core protein pV which can
CC be used in nucleic acid delivery vectors of the invention (the C-terminal
CC NLS being preferred).

XX Sequence 13 AA;

Query Match

Best Local Similarity 83.3%; Score 5; DB 22; Length 13;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KRVKK 6
| | | | |
Db 6 krvkk 10

RESULT 4

ID AAG57567 standard; Protein; 40 AA.

XX AAG57567;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 74200.

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140625.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

```

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145214.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0147038.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

```

```

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      83.3%; Score 5; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2 KRVKK 6
Db 15 KRVKK 19

```

```

RESULT 5
AAW08157
ID AAW08157 standard; Protein: 52 AA.
XX
AC AAW08157;
XX
DT 12-MAR-1997 (first entry)
XX
DE Drosophila PC class chromodomain.
XX
XX Bid; sex determination; chromodomain-Helicase-DNA binding 1;
KM CHD-1A; CHD-W; W chromosome; chromodomain; chromobox; PC.
XX
OS Drosophila sp.
XX
FH Key Location/Qualifiers
FT Domain 6..42
FT /label= Chromodomain
XX
PN W09639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
PI
PI Griffiths R, Tiwari B;
DR WPI; 1997-043127/04.
XX
PT Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
PS Disclosure; Fig 11; 76pp; English.
XX
CC The chromobox domain region (AAW08150) of the chicken sex determ.
CC CHD gene (see also AAT42751 and AAT42754) product was compared to
CC Chromodomain sequences from mouse CHD, yeast homologous CHD,

```

CC Drosophila HPI, human HPI, mouse MOD1 and MOD2, Drosophila PC and
 CC mouse MOD3 (AAW08151-58, respectively). The CHD class chromodomains
 CC included characteristic amino acid residues that were not found in
 CC the chromatin remodelling proteins not associated with sex determ.
 CC Proteins including these characteristic residues, when supplied to
 CC a cell or progenitor of an organism, activate the CHD-1A or
 CC CHD-W target gene thereby controlling the sex of the progeny.

XX Sequence 52 AA;

Query Match 83.3%; Score 5; DB 18; Length 52;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KRVKK 6

Db 14 KRVKK 18

RESULT 6

AAU00689

AAU00689 standard; Protein; 58 AA.

AAU00689;

07-SEP-2001 (first entry)

Thymosin-beta-10-1-like protein.

Thymosin-beta-10-1-like protein; ephrin type-A receptor 8-like protein;
 proteoglycan-like protein; fibromodulin; fibronectin; thymic immune cell;
 spermatogenesis; male infertility; neoplasia; red blood cell; platelet;
 small cell lung cancer; GPI-anchored ephrin-A ligand; prostate cancer;
 neurological disorder; cardiac disorder; vascular disorder; orthopaedic;
 inflammatory disease; rheumatoid arthritis; connective tissue;
 congenital muscular dystrophy; chemotherapy; immunotherapy.

Homo sapiens.

WO200129217-A2.

26-APR-2001.

13-OCT-2000; 2000WO-US28474.

15-OCT-1999; 99US-0159805.

18-OCT-1999; 99US-0159992.

22-OCT-1999; 99US-0160952.

12-OCT-2000; 2000US-0159805.

(CURA-) CURAGEN CORP.

Preyaga SK, Taupier RJ, Bandaru R;

WPI: 2001-308489/32.

N-PSDB: AAS00689.

New isolated polypeptides, NOV 1-3, having identity to
 thymosin-beta-10, ephrin type-A receptor 8 and proteoglycans, and
 polynucleotides, useful for treating male infertility, neurological or
 cardiac disease or rheumatoid arthritis -

Claim 1: Page 8; 102pp; English.

The sequence represents a thymosin-beta-10-1-like protein (NOV1).
 Thymosin-beta-10-1-like protein, ephrin type-A receptor 8-like protein
 (NOV2) and proteoglycan-like proteins (NOV3) may be used in the
 diagnosis, treatment and prevention of disorders caused by abnormal
 expression or activity of thymosin-beta-10, ephrin type-A receptor 8 and
 proteoglycans such as fibromodulin and fibronectin. The polypeptides of
 the invention are useful in screening for agents that modulate their
 activity, and in determining predispositions to disorders. NOV1 is useful

CC for treating conditions involving development, differentiation, and
 CC activation of thymic immune cells, in pathologies related to
 CC spermatogenesis and male infertility, diagnosis of neoplasias, in
 CC diseases or pathologies of red blood cells or platelets, in detection of
 CC small cell lung cancer. NOV1 nucleic acids can be combined in
 CC chemo-immunotherapeutic anti-cancer treatments. NOV2 is useful for
 CC detecting cells expressing GPI-anchored ephrin-A ligands, as a marker for
 CC prostate cancer, and in treating neurological, cardiac and vascular
 CC disorders. NOV3 (proteoglycan) nucleic acids and proteins are useful for
 CC treating orthopaedic disorders and/or injuries, and inflammatory diseases
 CC of connective tissues e.g. rheumatoid arthritis, congenital muscular
 CC dystrophies.

XX Sequence 58 AA;

Query Match 83.3%; Score 5; DB 22; Length 58;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ERVKK 5

Db 37 ERVKK 41

RESULT 7

AAW18344

AAW18344 standard; Protein; 70 AA.

AAW18344;

23-JUL-1997 (first entry)

Protein fragment encoded by bovine PIT-1 exon 6.
 Primer; polymerase chain reaction; PCR; amplify; detection; RFLP;
 HinfI polymorphism; exon 6; bovine; pituitary; PIT-1 gene; genotype;
 high milk production; restriction fragment length polymorphism.

Bos taurus.

US5614364-A.

25-MAR-1997.

16-MAY-1994; 94US-0243543.

16-MAY-1994; 94US-0243543.

(IOWA) UNIV IOWA STATE RES FOUND INC.

Freeman AE, Tuggle CK;

WPI: 1997-201471/18.

N-PSDB: AAT68932.

Identification of cows giving high milk yields - on basis of
 polymorphism in exon 6 of PIT-1 gene

Disclosure: Fig 4; 22pp; English.

This sequence is encoded by exon 6 of the bovine pituitary derived PIT-1
 gene, which has the HinfI polymorphic site. The PIT-1 DNA sequence was
 CC amplified using the primers given in AAT68930-31 for detecting the HinfI
 CC polymorphism. The method of the invention identifies a bovine with
 CC a genotype indicative of high milk production. The method comprises
 CC obtaining a nucleic acid sample from the bovine and ascertaining that
 CC the base at nucleotide position 165 in exon 6 of the bovine PIT-1 gene
 CC is adenine, or obtaining a sample of genomic DNA, digesting the sample
 CC with HinfI, separating the resulting fragments, and ascertaining that
 CC there is no HinfI site at nucleotide position 165 in exon 6 of the
 CC bovine PIT-1 gene. These primers are used to amplify the bovine PIT-1
 CC gene sequence, then restriction fragment length polymorphism analysis is

CC performed to detect an 800 bp fragment (indicating absence of a HinfI
CC site). Cattle with the genotype are selected for breeding. The genotype
CC lacking the HinfI site is associated with an average 13.2% increase in
CC milk production and an average 17.2% increase in total protein.
SQ Sequence 70 AA;

Query Match 83.3%; Score 5; DB 18; Length 70;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EKRK 5
Db 48 ekrvk 52

RESULT 8

AA017299 standard; Protein: 107 AA.

AA017299;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 18265.

Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125768.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156598.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 83.3%; Score 5; DB 21; Length 107;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KRYKK 6
DB 77 KRYKK 81
RESULT 9
AACG37567
ID AACG37567 standard; protein; 111 AA.
XX
XX AACG37567;
XX AC
XX DT 18-OCT-2000 (first entry)
XX DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 46210.
XX DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX PN
XX EP1033405-A2.
XX PD
XX 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0136847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 16-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 83.3%; Score 5; DB 21; Length 111;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KRYKK 6
|||||

Db 77 krchk 81

RESULT 10

AAg10185

XX AAg10185 standard; Protein: 121 AA.

AC AAg10185;

XX 17-OCT-2000 (first entry)

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 8403.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 8403.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0136229.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

```

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      83.3%; Score 5; DB 21; Length 121;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 KRYVK 6
    |||||
Db 58 kryvk 62

```

```

RESULT 11
AA01051
ID AA01051 standard; Protein; 132 AA.
XX

```

```

AC AA01051;
DT 06-OCT-2000 (first entry)
DE Human secreted protein, SEQ ID NO: 5132.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GIST ) GENSEPT.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX
XX N-PSDB; AAC01057.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 5132; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 132 AA;
SQ

```

```

Query Match      83.3%; Score 5; DB 21; Length 132;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 EKRVK 5
    |||||
Db 18 ekrvk 22

```

```

RESULT 12
AA017298
ID AA017298 standard; Protein; 140 AA.
XX
XX AA017298;
XX

```

```

DT 17-OCT-2000 (first entry)
XX

```

```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 18264.
XX

```

```

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

```

```

OS Arabidopsis thaliana.

```

XX EP103405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145319.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151330.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153707.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157111.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160776.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 83.3%; Score 5; DB 21; Length 140;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRVKK 6
Db 110 KRVKK 114

RESULT 13
AAG37566

ID AAG37566 standard; Protein; 144 AA.

AC AAG37566;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 46209.

KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0137725.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158332.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 83.3%; Score 5; DB 21; Length 144;
 Best Local Similarity 100.0%; Pred No. 47;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KRYKK 6
 |||||
 Db 110 krykk 114

RESULT 14
 AAB16726
 ID AAB16726 standard; Protein; 145 AA.
 XX
 AC AAB16726;
 XX
 DT 27-OCT-2000 (first entry)
 XX
 DE Bacteriophage Dp-1 protein sequence dp1ORF045.
 XX
 KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
 KW bacterial growth inhibition; bacterial infection.
 XX
 OS Bacteriophage Dp-1.
 XX
 PN WO200032825-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 03-DEC-1999; 99WO-IB02040.
 XX
 PR 03-DEC-1998; 98US-0110992.
 PR 03-JUN-1999; 99US-0326144.
 PR 28-SEP-1999; 99US-0407804.
 PR 30-SEP-1999; 99US-0157218.
 PR 01-DEC-1999; 99US-0168777.
 PR 02-DEC-1999; 99US-0454252.
 XX
 PA (PHAG-) PHAGETECH INC.

XX Pelletier J, Gros P, Dubow M;
XX WPI: 2000-412361/35.
DR N-PSDB: AAA69213.
XX
XX Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX
PS Example 17; Page 378; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAB68243 to AAB69442 and AAB16523 to AAB16554 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SQ Sequence 145 AA;

Query Match 83.38; Score 5; DB 21; Length 145;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KRVK 6
Db 2 KRVK 6

RESULT 15
AAG05046
ID AAG05046 standard; Protein: 148 AA.
XX
AC AAG05046;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1303.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 83.3%; Score 5; DB 21; Length 148;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KRVKK 6
|||
Db 73 KRVKK 77

Search completed: February 26, 2002, 01:50:47
Job time: 1251 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:45:56 ; Search time 78.18 Seconds
(without alignments)
11.226 Million cell updates/sec

Title: US-09-819-136-2_COPY_117_122
Perfect score: 6
Sequence: 1 RDRCEK 6

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	83.3	63	12 053548	053548 mycobacteri
2	5	83.3	86	12 099084	099084 hepatitis c
3	5	83.3	136	5 098139	098139 caenorhabdi
4	5	83.3	264	5 020043	020043 caenorhabdi
5	5	83.3	287	5 098139	098139 caenorhabdi
6	5	83.3	337	4 014868	014868 homo sapien
7	5	83.3	424	4 098139	098139 homo sapien
8	5	83.3	502	10 098139	098139 oryza sativ
9	5	83.3	555	10 098139	098139 arabidopsis
10	5	83.3	623	11 062510	062510 mus musculi
11	5	83.3	711	5 098139	098139 caenorhabdi
12	5	83.3	996	5 098139	098139 caenorhabdi
13	5	83.3	1041	5 023466	023466 caenorhabdi
14	5	83.3	2146	5 098139	098139 caenorhabdi
15	5	83.3	2879	5 098139	098139 caenorhabdi
16	4	66.7	42	8 032982	032982 plus thund
17	4	66.7	42	8 032982	032982 plus thund
18	4	66.7	51	8 032994	032994 plus thund
19	4	66.7	52	5 098139	098139 caenorhabdi

20	4	66.7	58	5 098139	098139 caenorhabdi
21	4	66.7	58	5 098139	098139 caenorhabdi
22	4	66.7	59	5 098139	098139 caenorhabdi
23	4	66.7	59	12 066082	066082 canine hept
24	4	66.7	61	2 098139	098139 streptococc
25	4	66.7	66	2 098139	098139 bruceella ab
26	4	66.7	68	2 098139	098139 bruceella ab
27	4	66.7	72	12 057256	057256 vaccinia vl
28	4	66.7	73	2 098139	098139 streptomyc
29	4	66.7	76	5 098139	098139 mycobacteri
30	4	66.7	83	2 045332	045332 bruceella ov
31	4	66.7	86	12 098139	098139 human immun
32	4	66.7	88	12 098139	098139 human immun
33	4	66.7	88	12 098139	098139 human immun
34	4	66.7	89	1 026398	026398 methanobact
35	4	66.7	89	2 044071	044071 aeromonas h
36	4	66.7	89	10 098139	098139 oryza sativ
37	4	66.7	90	1 026580	026580 methanobact
38	4	66.7	94	2 098139	098139 papaya bunc
39	4	66.7	94	2 098139	098139 burkholderi
40	4	66.7	97	2 098139	098139 streptomyc
41	4	66.7	100	1 027828	027828 methanobact
42	4	66.7	101	5 098139	098139 dirosophila
43	4	66.7	103	3 098139	098139 neurospora
44	4	66.7	108	3 002255	002255 pneumocysti
45	4	66.7	108	5 045455	045455 caenorhabdi

ALIGNMENTS

RESULT 1					
ID	053548	PRELIMINARY:	PRT:	63 AA.	
AC	053548:				
DT	01-JUN-1998 (TREMBlrel. 06, Created)				
DT	01-JUN-1998 (TREMBlrel. 06, Last sequence update)				
DE	01-JUN-2000 (TREMBlrel. 14, Last annotation update)				
DE	POTRATIVE FERREDOXIN.				
GN	FDX OR RV3503C OR MTV023.10C.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsbly T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Sultun J.E.,				
RA	Taylor K., Whitehead S., Barrell B.G.,				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
DR	EMBL; AL022022; CAAL17740.1;				
DR	HSSP; P46797; 1VTM.				
DR	Tuberculin; RV3503C;				
KW	Complete proteome.				
SC	SEQUENCE 63 AA; 6904 MW; 2C29D2CE701421E0 CRC64;				
Query Match	83.3%;	Score 5;	DB 2;	Length 63;	
Best Local Similarity	100.0%;	Pred. No. 11;			
Matches 5;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1 RDRCE 5				
DB					
	7 RDRCE 11				

```

RESULT 2
Q990R4 PRELIMINARY; PRT; 86 AA.
AC Q990R4;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCP54;
RA Pulg-Basagoliti F., Saiz J.C., Fornis X., Ampurdanes S.,
RA Gimenez-Barcons M., Franco S., Sanchez-Fueyo A., Costa J.,
RA Sanchez-Tapias J.M., Rodes J.;
RT "Influence of the genetic heterogeneity of the ISDR and NS5B regions
RT of hepatitis C virus on the response to interferon therapy in chronic
RT hepatitis C.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL_A0339317; AAK19729.1; -.
FT NON_TER. 86
FT 1
SQ SEQUENCE 86 AA; 9616 MW; 039A316A853229A7 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 12; Length 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DRCEK 5
Db 1 DRCEK 5
|||||

```

```

Query Match
Best Local Similarity 83.3%; Score 5; DB 5; Length 136;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRCEK 6
|||||

```

```

Db 71 DRCEK 75

RESULT 4
Q20043 PRELIMINARY; PRT; 264 AA.
AC Q20043;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SIMILAR TO EGF-LIKE REPEATS.
GN F35D2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Waterson R., Watson A., Watson K., Vaudin M., Vaughan K., Waterson R.,
RA Waterson R., Watson K., Vaudin M., Vaughan K., Waterson R.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Connell M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterson R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL_U28741; AAA68325.1; -.
DR HSSP; P04070; 1AUF.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR SMART; SM00180; EGF_Lam; 1.
DR SMART; SM00001; EGF_Like; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 264 AA; 28434 MW; C91B1CC7B4B463F4 CRC64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 5; Length 264;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRCEK 6
Db 83 DRCEK 87
|||||

```

```

RESULT 5
Q9GY27 PRELIMINARY; PRT; 287 AA.
AC Q9GY27;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE POSSIBLE PUTATIVE MAGNESIUM AND COBALT TRANSPORT PROTEIN (FRAGMENT).

```

GN LM12.131.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL30114; CAC02023.1; -
DR InterPro: IPR002523; CORA.
DR Pfam: PF01544; CORA; 1.
FT NON_TER 1 1
FT SEQUENCE 287 287
SO SEQUENCE 287 AA; 32353 MW; 476B26A8991AF98A CRC64;

Query Match 83.3%; Score 5; DB 5; Length 287;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRCE 5
DB 125 RDRCE 129

RESULT 6
ID 014868 PRELIMINARY; PRT; 337 AA.
AC 014868;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE EFFECTOR CELL PROTEASE RECEPTOR 1 (FRAGMENT).
GN EPR-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94148797; PubMed=8106347;
RA Altiert D.C.;
RT "Molecular cloning of effector cell protease receptor-1, a novel cell
RT surface receptor for the protease factor Xa.";
RL J. Biol. Chem. 269:3139-3142(1994).
DR EMBL: L26245; AAA19687.1; -
KW Protease.
FT NON_TER 337 337
SO SEQUENCE 337 AA; 36825 MW; 72B5AF24E8C8C09F CRC64;

Query Match 83.3%; Score 5; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRCE 5
DB 58 RDRCE 62

RESULT 7
ID 09NS19 PRELIMINARY; PRT; 424 AA.
AC 09NS19;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PRD4 PROTEIN (FRAGMENT).
GN PRD4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hattori M., Fujiyama A., Taylor T.D., Matanabe H., Yada T., Park H.S.,
RA Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillnabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloedeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Dagand E.,
RA Wehmeyer S., Bozzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.L.;
DR EMBL: AL163203; CAB90394.1; -
FT NON_TER 1 1
FT SEQUENCE 424 424
SO SEQUENCE 424 AA; 50038 MW; C6E704795C534625 CRC64;

Query Match 83.3%; Score 5; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRCE 5
DB 317 RDRCE 321

RESULT 8
ID 09LH28 PRELIMINARY; PRT; 502 AA.
AC 09LH28;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILAR TO ARABIDOPSIS THALIANA DNA CHROMOSOME 2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriaridoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT clone:P0029D06.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001552; BAA33026.1; -
DR InterPro: IPR001440; TPR.
SO SEQUENCE 502 AA; 55490 MW; 276AB8BD13250F55E CRC64;

Query Match 83.3%; Score 5; DB 10; Length 502;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCEK 6
DB 298 DRCEK 302

RESULT 9
ID 09C7V7 PRELIMINARY; PRT; 555 AA.
AC 09C7V7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HYPOTHETICAL 64.5 KDA PROTEIN.
 GN F15H21.4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chen A., Chen Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizart L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
 RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana."
 RL Nature 408:816-820(2000).
 DR EMBL: AC066889; AAC51714.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 555 AA; 64494 MW; 3530A22BF67E23 CRC64;

Query Match 83.3%; Score 5; DB 10; Length 555;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DRCE 5
 DB 530 DRCE 534

RESULT 10
 ID 062510 PRELIMINARY; PRT; 623 AA.
 AC 062510;
 DT 01-NOV-1996 (TREMURel. 01, Created)
 DT 01-NOV-1996 (TREMURel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMURel. 17, Last annotation update)
 DE ZINC FINGER PROTEIN 62 (FRAGMENT).
 GN ZFP62 OR ZFP OR ZT3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EYE LENS;
 RA Brady J.F.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 102-623 FROM N.A.
 RC TISSUE=SKETAL MUSCLE;
 RX MEDLINE=96404132; PubMed=8808410;
 RA Polimeni M., Giorgi S., De Gregorio L., Dragani T.A., Molinaro M.,
 RA Cosu G., Bouche M.;
 RT "Differentiation dependent expression in muscle cells of ZT3, a novel
 RT zinc finger factor differentially expressed in embryonic and adult
 RT tissues."
 RL Mech. Dev. 54:107-117(1996).
 DR EMBL: L36315; AAA40580.1; -.

DR EMBL: Z67747; CAA91560.1; -.
 DR HSSP: P08047; 1SP2.
 DR MGD: MGI:99662; Zfp62.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF00096; Zf-C2H2; 19.
 DR SMART: SM00355; Znf-C2H2; 19.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 19.
 FT DNA-binding; Metal-binding; Zinc-finger.
 FT NON_TER
 SQ SEQUENCE 623 AA; 71434 MW; 3F56836AC2FB3BD0 CRC64;

Query Match 83.3%; Score 5; DB 11; Length 623;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCE 6
 DB 347 DRCE 351

RESULT 11
 ID 09XWC4 PRELIMINARY; PRT; 711 AA.
 AC 09XWC4;
 DT 01-NOV-1999 (TREMURel. 12, Created)
 DT 01-NOV-1999 (TREMURel. 12, Last sequence update)
 DE 01-JUN-2001 (TREMURel. 17, Last annotation update)
 DE Y70G10A.2 PROTEIN.
 GN Y70G10A.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd C.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ritken L., Koopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 RL EMBL: AL032660; CAA21752.1; -.
 DR InterPro: IPR000561; EGF-like.
 DR SMART: SM00181; EGF; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 2.
 KW EGF-like domain; Glycoprotein.
 SQ SEQUENCE 711 AA; 81241 MW; C12DE08BDA1CD116 CRC64;

Query Match 83.3%; Score 5; DB 5; Length 711;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCE 6
 DB 197 DRCE 201

RESULT 12
Q9VU34 PRELIMINARY: PRT: 996 AA.
AC Q9VU34.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG11259 PROTEIN.
GN CG11259.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner S.N., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bendale J., Braykaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brooks-Taylor P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu L., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foster C., Garfield A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwu C.,
RA Jabali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svlerks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC EMBL: AE003539; AAF49857.1; -.
CC HSP: 001082; 1BKR.
CC PDBase: PDB0036333; CG11259.
DR InterPro: IPR001715; Calponin_hom.
DR InterPro: IPR01781; LIM.
DR Pfam: PF00307; CH; 1.
DR SMART: SM00412; LIM; 1.
DR SMART: SM00033; CH; 1.
DR SMART: SM00132; LIM; 1.
DR PROSITE: PS50021; CH; 1.
DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
DR LIM motif: Metal-binding; ZINC.
KW SEQUENCE 996 AA; 112194 MW; DB9AD02DE81C01EE CRC64;

Query Match 83.3%; Score 5; DB 5; Length 996;
Best local similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RDCE 5
Db 794 RDCE 798

RESULT 13
Q23466 PRELIMINARY: PRT: 1041 AA.
ID Q23466
AC Q23466;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILAR TO UBIQUITIN CONJUGATING ENZYME.
GN ZK328.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
RA Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopa A., Saunders R., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sultson J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A.;
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO EF-HAND FAMILY.
CC EMBL: U50193; AAA91251.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001394; UCH-2.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00442; UCH-1; 1.
DR Pfam: PF00443; UCH-2; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS50235; UCH_2_3; 1.
KW Calcium-binding
SQ SEQUENCE 1041 AA; 118414 MW; 6D8991090FC7F3C6 CRC64;

Query Match 83.3%; Score 5; DB 5; Length 1041;
Best local similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DRCEK 6
Db 916 DRCEK 920

RESULT 14
Q9VC97 PRELIMINARY: PRT: 2146 AA.
ID Q9VC97;
AC Q9VC97;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CRB PROTEIN.
 GN CRB OR CG383.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amentides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis J.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Kalush H., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush H., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Maasman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003747; AAF56276.1; -.
 DR HSSP: P00740; IIXA.
 DR FLYBASE: Fbgn0000368; crb.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam: PF00008; EGF; 27.
 DR Pfam: PF00054; laminin_G; 3.
 DR PRINTS: PRO0010; EGFBL00D.
 DR PRINTS: PRO0011; EGF_LAMININ.
 DR SMART: SM00179; EGF_CA; 11.
 DR SMART: SM00001; EGF_like; 16.
 DR SMART: SM00282; LamG; 3.
 DR PROSITE: PS00010; ASX_HYDROXYL; 16.
 DR PROSITE: PS00022; EGF_1; 25.
 DR PROSITE: PS01186; EGF_2; 17.
 DR PROSITE: PS01187; EGF_CA; 13.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 2146 AA; 233570 MW; 8E23B9E32B76115 CRC64;

Query Match 83.3%; Score 5; DB 5; Length 2146;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DRCEK 6
 DB 576 DRCEK 580
 RESULT 15
 ID Q9U0Z1 PRELIMINARY; PRT; 2879 AA.
 AC Q9U0Z1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 305.0 KDA PROTEIN.
 GN U5883.06.
 OS Leishmania major.
 OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Murphy L., Harris D., Ivens A.C., Lawson D., Quail M.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL: AL117384; CAB5617.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 2879 AA; 305039 MW; 1F75F7831C99B7DB CRC64;

Query Match 83.3%; Score 5; DB 5; Length 2879;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DRCEK 6
 DB 2551 DRCEK 2555

Search completed: February 26, 2002, 01:53:16
 Job time: 440 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:46:51 ; Search time 24.25 Seconds

Title: US-09-819-136-2_COPY_117_122

```
Perfect score: 6
Sequence: 1 RDRCEK 6
```

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

Post-processing: Listing first 45 summaries

Database : SwlssProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5	83.3	183	1	LIPO_BUTMA	Q01584 bufo marinus
2	5	83.3	236	1	TL17_ARATH	P01160 arabidopsis
3	5	83.3	964	1	UL70_MCVWS	O69153 murine cyto
4	5	83.3	2139	1	CRB_DROME	P10040 drosophila
5	4	66.7	23	1	TL17_SPTOL	P81778 sphinctra ol
6	4	66.7	62	1	IP52_ANESU	P10280 anemonia su
7	4	66.7	65	1	ICAA_ASCSU	P19399 ascaris suu
8	4	66.7	77	1	VB09_VACCOC	P21005 vaccinia v
9	4	66.7	85	1	COXG_BOVIN	P00429 bos taurus
10	4	66.7	85	1	COXG_MOUSE	P56391 mus musculu
11	4	66.7	85	1	RPOL_METHH	O27372 methanobact
12	4	66.7	94	1	Y172_BURCE	P24578 burkholderi
13	4	66.7	105	1	MIH_METEN	O76534 metapneus
14	4	66.7	109	1	TRP5_BOVIN	O9my9v bos taurus
15	4	66.7	112	1	VG33_BP14	P13338 bacteriophag
16	4	66.7	114	1	PEG6_YEAST	P52553 saccharomyce
17	4	66.7	118	1	Y580_MERTJA	O58000 methanococc
18	4	66.7	127	1	PA2H_XENELA	P41485 xenopus lae
19	4	66.7	128	1	LYC1_PIG	P12057 sus scrofa
20	4	66.7	143	1	LCA_MOUSE	P29792 mus musculu
21	4	66.7	144	1	LYC2_ONCMY	P11941 oncorhynchus
22	4	66.7	145	1	IA01_MHEAT	P16850 triticum ae
23	4	66.7	145	1	IA02_MHEAT	P16851 triticum ae
24	4	66.7	146	1	LYC3_PIG	P12068 sus scrofa
25	4	66.7	148	1	LYC3_PIG	P12069 sus scrofa
26	4	66.7	153	1	Y063_NPVAC	P41466 autographa
27	4	66.7	156	1	IF2B_ARATH	Q041969 arabidopsis
28	4	66.7	159	1	LCA_RAT	P00714 rattus norv
29	4	66.7	175	1	RP18_MOUSE	P47939 mus musculu
30	4	66.7	175	1	RP18_RAT	P47940 rattus norv
31	4	66.7	183	1	CAL3_PETHY	P27154 petunia hyb
32	4	66.7	183	1	VNS2_MOUSE	O62472 mus musculu
33	4	66.7	185	1	CASK_BOVIN	P02668 bos taurus

34	4	66.7	190	1	CASK_BOBBU	P18480 bubabu bub
35	4	66.7	194	1	V182_METU	O57641 methanoccc
36	4	66.7	197	1	T4S5_HUMAN	O14869 homo sapien
37	4	66.7	199	1	YCFP_ECOLI	P75950 escherichia
38	4	66.7	203	1	ET1_PIG	P09558 sus scrofa
39	4	66.7	203	1	HM16_CAEBL	P34326 caenorhabdi
40	4	66.7	204	1	SRY_PANPA	O28778 pan paniscu
41	4	66.7	214	1	VEFR_PSEAE	P55222 pseudomona
42	4	66.7	216	1	VEGF_CHICK	P52582 gallus gall
43	4	66.7	224	1	YF57_MYCTU	O50741 mycobacteri
44	4	66.7	225	1	DL04_HSVEB	P28943 equine herp
45	4	66.7	227	1	SRY_CALJA	P51501 callithrix

ALIGNMENTS

RESULT	1
LIPO_BUFMA	
ID	LIPO_BUFMA
AC	001584;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-APR-1993 (Rel. 25, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	LIPOCALIN PRECURSOR.
OS	Bufo marinus (Giant toad) (Cane toad).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae;
OC	Bufo.
ON	NCBI_Taxid=8386;
RP	[1]
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-16; 18-20; 67-73; 86-93;136-141.
RC	TISSUE=Choroid plexus;
RX	MEDLINE=93054646; PubMed=1385415;
RA	Achen M.G., Harns P.J., Thomas T., Richardson S.J.,
RA	Wettenhall R.E.H., Schneider G.,
RT	"Protein synthesis at the blood-brain barrier. The major protein
RT	secreted by amphibian choroid plexus is a lipocalin."
RL	J. Biol. Chem. 267:23170-23174(1992).
CC	-1- FUNCTION: MIGHT HAVE A TRANSPORT FUNCTION ACROSS THE BLOOD BRAIN
CC	BARRIER. IS SUPPOSED TO HAVE SIMILAR FUNCTIONS AS TRANSHYRETIN
CC	WHICH MUST HAVE EVOLVED AFTER THE STAGE OF THE AMPHIBIANS IN
CC	EVOLUTION.
CC	-1- SUBUNIT: MONOMER.
CC	-1- SUBCELLULAR LOCATION: SECRETED INTO THE CEREBROSPINAL FLUID.
CC	-1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN CHOROID PLEXUS. MUCH LOWER
CC	EXPRESSION IN OTHER BRAIN AREAS, AND ABSENT FROM LIVER.
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT AMPHIBIAN METAMORPHOSIS.
CC	-1- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on ways
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X67952; CAA48138.1; -;
DR	EMBL; L06806; AAA48554.1; -;
DR	PIR; S25465; S25465.
DR	PIR; A44456; A44456.
DR	InterPro; IPR002345; LIPOCALIN.
DR	InterPro; IPR000566; LIPOCALIN_CYTRBP.
DR	Pfam; PF00061; LIPOCALIN.1.
DR	PRINTS; PR00179; LIPOCALIN.
DR	PRINTS; PR01254; PGSDSYNTHASE.
DR	PROSITE; PS00213; LIPOCALIN; 1.
KW	LIPOCALIN; Signal.
FT	SIGNAL
FT	CHAIN
FT	DISULFID
SQ	SEQUENCE

Query Match 83.3%; Score 5; DB 1; Length 183;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCEK 6
 DB 81 DRCEK 85

RESULT 2

TL17_ARATH
 ID TL17_ARATH STANDARD: PRT: 236 AA.
 AC P81760;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THYLAKOID LUMENAL 17.4 KDA PROTEIN, CHLOROPLAST PRECURSOR (P17.4).
 GN ATG53490 OR MNC6.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98316713; PubMed=9654141;
 RA Kieselbach T., Mant A., Robinson C., Schroeder W.P.;
 RT "Characterization of an Arabidopsis cDNA encoding a thylakoid lumen
 RT protein related to novel 'pentapeptide repeat' family of proteins.";
 RL FEBS Lett. 428:241-244(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=99087489; PubMed=9872454;
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyaama N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:297-308(1998).
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; WITHIN THE THYLAKOID LUMEN.
 CC -1- SIMILARITY: CONTAINS 2 PENTAPEPTIDE REPEAT DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB015476; BAB09725.1;
 DR InterPro; IPR001646; Pentapeptide.
 DR Pfam; PF00805; Pentapeptide; 2.
 KW Chloroplast; Thylakoid; Transit peptide; Repeat.
 FT TRANSIT 1
 FT TRANSIT ?
 FT CHAIN 78 236 THYLAKOID (BY SIMILARITY).
 FT DOMAIN 124 163 THYLAKOID LUMENAL 17.4 KDA PROTEIN.
 FT DOMAIN 169 208 PENTAPEPTIDE 1.
 FT CONFLICT 21 21 PENTAPEPTIDE 2.
 FT CONFLICT 21 21 K -> N (IN REF. 2).
 SQ SEQUENCE 236 AA; 25658 MW; 234277E4652F1A47 CRC64;

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB015476; BAB09725.1;
 DR InterPro; IPR001646; Pentapeptide.
 DR Pfam; PF00805; Pentapeptide; 2.
 KW Chloroplast; Thylakoid; Transit peptide; Repeat.
 FT TRANSIT 1
 FT TRANSIT ?
 FT CHAIN 78 236 THYLAKOID (BY SIMILARITY).
 FT DOMAIN 124 163 THYLAKOID LUMENAL 17.4 KDA PROTEIN.
 FT DOMAIN 169 208 PENTAPEPTIDE 1.
 FT CONFLICT 21 21 PENTAPEPTIDE 2.
 FT CONFLICT 21 21 K -> N (IN REF. 2).
 SQ SEQUENCE 236 AA; 25658 MW; 234277E4652F1A47 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 236;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCEK 6
 DB 90 DRCEK 94

RESULT 3
 UL70_MCMVS
 ID UL70_MCMVS STANDARD: PRT: 964 AA.
 AC Q69153;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HELICASE/PRIMASE COMPLEX PROTEIN (PROBABLE DNA REPLICATION PROTEIN
 DE UL70).
 GN UL70.
 OS Murine cytomegalovirus (strain Smith).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Muromegalovirus.
 OX NCBI_TaxId=10367;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96082764; PubMed=7483291;
 RA Messerle M., Rapp M., Lucin P., Koszinowski U.H.;
 RT "Characterization of a conserved gene block in the murine
 RT cytomegalovirus genome.";
 RL Virus Genes 10:73-80(1995).
 CC -1- FUNCTION: INVOLVED IN DNA REPLICATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL52,
 CC EBV-1 7, EBV BSLF1, HVS-1 56, VZV 6, HCMV AND MCMV UL70.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L07319; AAA96665.1;
 KW Helicase; DNA replication.
 SQ SEQUENCE 964 AA; 109704 MW; 70605300E1D85864 CRC64;

Query Match 83.3%; Score 5; DB 1; Length 964;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DRCEK 5
 DB 238 DRCEK 242

RESULT 4

CRB_DROME
 ID CRB_DROME STANDARD: PRT: 2139 AA.
 AC P10040;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CRUMBS PROTEIN PRECURSOR (95F).
 GN CRB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORECON-R; TISSUE=Embryo;
 RX MEDLINE=90263104; PubMed=2344615;
 RA Tepass U., Theres C., Knust E.;
 RT "Crumbs encodes an EGF-like protein expressed on apical membranes of
 RT Drosophila epithelial cells and required for organization of
 RT epithelia.";
 RL Cell 61:787-799(1990).
 RN [2]

RP SEQUENCE OF 1663-1955 FROM N.A.
RC TISSUE-Embryo;
RX MEDLINE-87218537; PubMed-3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,
RA Vaesslin H., Campos-Ortega J.A.;
RT "EGF homologous sequences encoded in the genome of Drosophila
RT melanogaster, and their relation to neurogenic genes.";
RL EMBL J. 6:761-766(1987).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC POLARITY. IT MAY ACT AS A SIGNAL.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PFM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CA28793.1; -.
DR PIR; B26637; B26637.
DR PIR; A35672; A35672.
DR HSSP; P00740; 11XA.
DR FlyBase; FBgn0000368; crb.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF_Like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_IT.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 27.
DR PRINTS; PR00010; EGFBL0D.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00001; EGF_Like; 16.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 15.
KW Differentiation; Repeat; EGF-Like domain; Transmembrane;
KW Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 90
FT CHAIN 1 2139 CRUMBS PROTEIN.
FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303 EGF-Like 1.
FT DOMAIN 306 343 EGF-Like 2.
FT DOMAIN 348 386 EGF-Like 3.
FT DOMAIN 388 425 EGF-Like 4.
FT DOMAIN 427 463 EGF-Like 5.
FT DOMAIN 464 500 EGF-Like 6.
FT DOMAIN 501 532 EGF-Like 7.
FT DOMAIN 545 581 EGF-Like 8.
FT DOMAIN 582 611 EGF-Like 9.
FT DOMAIN 609 646 EGF-Like 10.
FT DOMAIN 648 685 EGF-Like 11.
FT DOMAIN 687 723 EGF-Like 12.
FT DOMAIN 725 760 EGF-Like 13.
FT DOMAIN 763 801 EGF-Like 14.
FT DOMAIN 802 838 EGF-Like 15.
FT DOMAIN 840 902 EGF-Like 16.
FT DOMAIN 904 940 EGF-Like 17.
FT DOMAIN 942 978 EGF-Like 18.
FT DOMAIN 980 1021 EGF-Like 19.
FT DOMAIN 1207 1243 EGF-Like 20.
FT DOMAIN 1481 1517 EGF-Like 21.

FT DOMAIN 1759 1795 EGF-Like 22.
FT DOMAIN 1797 1833 EGF-Like 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1835 1871 EGF-Like 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1874 1915 EGF-Like 25.
FT DOMAIN 1915 1951 EGF-Like 26.
FT DOMAIN 1953 1989 EGF-Like 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1991 2029 EGF-Like 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2030 2070 EGF-Like 29.
FT DISULFD 271 282 BY SIMILARITY.
FT DISULFD 276 291 BY SIMILARITY.
FT DISULFD 293 302 BY SIMILARITY.
FT DISULFD 310 321 BY SIMILARITY.
FT DISULFD 315 331 BY SIMILARITY.
FT DISULFD 333 342 BY SIMILARITY.
FT DISULFD 352 363 BY SIMILARITY.
FT DISULFD 357 374 BY SIMILARITY.
FT DISULFD 376 385 BY SIMILARITY.
FT DISULFD 392 403 BY SIMILARITY.
FT DISULFD 397 412 BY SIMILARITY.
FT DISULFD 414 424 BY SIMILARITY.
FT DISULFD 431 442 BY SIMILARITY.
FT DISULFD 436 451 BY SIMILARITY.
FT DISULFD 453 462 BY SIMILARITY.
FT DISULFD 468 479 BY SIMILARITY.
FT DISULFD 473 488 BY SIMILARITY.
FT DISULFD 490 499 BY SIMILARITY.
FT DISULFD 505 515 BY SIMILARITY.
FT DISULFD 509 520 BY SIMILARITY.
FT DISULFD 522 531 BY SIMILARITY.
FT DISULFD 549 562 BY SIMILARITY.
FT DISULFD 556 569 BY SIMILARITY.
FT DISULFD 571 580 BY SIMILARITY.
FT DISULFD 586 597 BY SIMILARITY.
FT DISULFD 591 602 BY SIMILARITY.
FT DISULFD 604 610 BY SIMILARITY.
FT DISULFD 613 624 BY SIMILARITY.
FT DISULFD 618 634 BY SIMILARITY.
FT DISULFD 636 645 BY SIMILARITY.
FT DISULFD 652 664 BY SIMILARITY.
FT DISULFD 659 673 BY SIMILARITY.
FT DISULFD 675 684 BY SIMILARITY.
FT DISULFD 691 702 BY SIMILARITY.
FT DISULFD 696 711 BY SIMILARITY.
FT DISULFD 713 722 BY SIMILARITY.
FT DISULFD 729 740 BY SIMILARITY.
FT DISULFD 734 749 BY SIMILARITY.
FT DISULFD 751 760 BY SIMILARITY.
FT DISULFD 767 778 BY SIMILARITY.
FT DISULFD 772 787 BY SIMILARITY.
FT DISULFD 789 799 BY SIMILARITY.
FT DISULFD 806 817 BY SIMILARITY.
FT DISULFD 811 826 BY SIMILARITY.
FT DISULFD 828 837 BY SIMILARITY.
FT DISULFD 844 855 BY SIMILARITY.
FT DISULFD 849 890 BY SIMILARITY.
FT DISULFD 892 901 BY SIMILARITY.
FT DISULFD 908 919 BY SIMILARITY.
FT DISULFD 913 928 BY SIMILARITY.
FT DISULFD 930 939 BY SIMILARITY.
FT DISULFD 946 957 BY SIMILARITY.
FT DISULFD 952 966 BY SIMILARITY.
FT DISULFD 968 977 BY SIMILARITY.
FT DISULFD 984 995 BY SIMILARITY.
FT DISULFD 989 1009 BY SIMILARITY.
FT DISULFD 1011 1022 BY SIMILARITY.
FT DISULFD 1211 1220 BY SIMILARITY.
FT DISULFD 1216 1231 BY SIMILARITY.
FT DISULFD 1233 1242 BY SIMILARITY.
FT DISULFD 1485 1496 BY SIMILARITY.
FT DISULFD 1490 1505 BY SIMILARITY.
FT DISULFD 1507 1516 BY SIMILARITY.
FT DISULFD 1763 1774 BY SIMILARITY.
FT DISULFD 1768 1783 BY SIMILARITY.

```

FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 565 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 736 736 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 884 884 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match
Best Local Similarity 100.0%; Score 5; DB 1; Length 2139;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCE 6
Db 578 DRCEK 582

```

RESULT 5
TL17_SPTOL STANDARD; PRT; 23 AA.
ID TL17_SPTOL
AC P81778;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THYLAKOID LUMENAL 17.4 KDA PROTEIN (P17.4) (FRAGMENT).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN 11
RP SEQUENCE AND REVISIONS TO 13 AND 15.
RA Schubert M., Peterson U., Funk C., Schroeder W.P., Kieselbach T.;
RT "The chloroplast lumen from Arabidopsis thaliana."
RL Submitted (JUL-2001) to the SWISS-PROT data bank.
RN 121
RP SEQUENCE OF 1-16.
RC TISSUE=Leaf;
RX MEDLINE=98175931; PubMed=9506969;
RA Kieselbach T., Hagman A., Anderson B., Schroeder W.P.;
RT "The thylakoid lumen of chloroplasts - isolation and
characterization."
RL J. Biol. Chem. 273:6710-6716(1998).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; WITHIN THE THYLAKOID LUMEN.
KW Chloroplast; Thylakoid.
FT NON_TER 23
SQ SEQUENCE 23 AA; 2571 MW; 022C92717AE9AF2D CRC64;

```

Query Match
Best Local Similarity 100.0%; Score 4; DB 1; Length 23;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCE 5
Db 13 DRCE 16

```

RESULT 6
IP52_ANESU STANDARD; PRT; 62 AA.
ID IP52_ANESU
AC P10280;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEASE INHIBITOR 5 II (SAs II).
DE Anemonia sulcata (Snake-locks sea anemone).
OS Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
OC Nymantidae; Actinidae; Anemonia.
OX NCBI_TaxID=6108;
RN 111
RP SEQUENCE OF 1-59.
RA Wunderer G., Machleidt W., Fritz H.;
RT "The broad-specificity proteinase inhibitor 5 II from the sea anemone
Anemonia sulcata."
RL Meth. Enzymol. 80:816-820(1981).
RN 121
RP SEQUENCE.
RA Krebs H.C., Habermehl G.G.;
RT "Isolation and structural determination of a hemolytic active peptide
from the sea anemone Metridium senile."
RL Naturwissenschaften 74:395-396(1987).
CC -1- FUNCTION: INHIBITOR OF KALLIKREINS.
CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
DR PIR; B27222; B27222.
DR PIR; S07451; S07451.
DR HSSP; P31713; ISHP.
DR InterPro: IPR002223; Kunitz_BPTI.
DR Pfam: PF000014; Kunitz_BPTI; 1.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00131; KU; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor.
FT DISULFID 5 55 BY SIMILARITY.
FT DISULFID 14 38 BY SIMILARITY.
FT DISULFID 30 51 BY SIMILARITY.
FT ACT_SITE 15 16 REACTIVE BOND (BY SIMILARITY).
FT VARIANT 13 13 P -> R.
FT VARIANT 16 16 A -> G.
FT VARIANT 17 17 R -> G.
FT VARIANT 25 25 S -> L.
FT VARIANT 28 28 K -> R.
FT VARIANT 39 39 G -> R.
SQ SEQUENCE 62 AA; 6937 MW; 7262D028CA567BC8 CRC64;

```

Query Match
Best Local Similarity 100.0%; Score 4; DB 1; Length 62;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
Db 29 RCEK 32

```

RESULT 7
ICAA_ASCSU STANDARD; PRT; 65 AA.
ID ICAA_ASCSU
AC P19399;
DT 01-NOV-1990 (Rel. 16, Created)

```

DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DE 01-FEB-1996 (Rel. 33, Last annotation update)
 OS CARBOXYPEPTIDASE A INHIBITOR.
 CC Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 CC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=89192365; PubMed=2648995;
 RA Homandberg G.A., Lilwaller R.D., Peanasky R.J.;
 RT "Carboxypeptidase inhibitors from Ascaris suum: the primary
 structure."
 RL Arch. Biochem. Biophys. 270:153-161(1989).
 DR PIR. S03858; S03858.
 KW Protease inhibitor.
 SQ SEQUENCE 65 AA; 7511 MW; E84DC9CA8326180B CRC64;

Query Match 66.7%; Score 4; DB 1; Length 65;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 DB 33. RCEK 36

RESULT 8
 VE09_VACCC STANDARD; PRT; 77 AA.
 AC P21005;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN B9 PRECURSOR.
 GN B9.
 OS Vaccinia virus (strain Copenhagen).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10249;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkus M.E., Winslow J.P.,
 RA Paolletti E.;
 RT "The complete DNA sequence of vaccinia virus."
 RL Virology 179:247-266(1990).
 RN [2]
 RP COMPLETE GENOME.
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RT "Appendix to 'The complete DNA sequence of vaccinia virus'."
 RL Virology 179:517-563(1990).
 CC -1- SIMILARITY: TO SHOPE FIBROMA VIRUS AND CAPRIPOXVIRUS (STRAIN
 INS-1) T4 PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M35027; AAA48206.1; -.
 DR PIR. H42526; H42526.
 KW Signal.
 FT SIGNAL. 1 17 POTENTIAL.
 FT CHAIN 18 77 PROTEIN B9.
 FT SEQUENCE 77 AA; 8806 MW; C60986C87B0AC9BB CRC64;

Query Match 66.7%; Score 4; DB 1; Length 77;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 DB 21 RCEK 24

RESULT 9
 COXG_BOVIN STANDARD; PRT; 85 AA.
 ID COXG_BOVIN
 AC P00429;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE VIB (EC 1.9.3.1) (AED).
 GN COX6B.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=80070615; PubMed=229067;
 RA Steffens G.C.M., Steffens G.J., Buse G.;
 RT "Studies on cytochrome c oxidase, VIII. The amino acid sequence of
 RT polypeptide VII."
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:1641-1650(1979).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81191870; PubMed=6262305;
 RA Tanaka M., Yasunobu K.T., Wei Y.-H., King T.E.;
 RT "The complete amino acid sequence of bovine heart cytochrome oxidase
 RT subunit VI."
 RL J. Biol. Chem. 256:4832-4837(1981).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=89345180; PubMed=2548168;
 RA Lightowers R.N., Capaldi R.A.;
 RT "Nucleotide sequence of the cDNA encoding subunit AED (VIB) of beef
 RT heart cytochrome c oxidase."
 RL Nucleic Acids Res. 17:5845-5845(1989).
 RN [4]
 RP SEQUENCE OF 44-66.
 RC TISSUE=Liver;
 RX MEDLINE=89000697; PubMed=2844245;
 RA Yanamura W., Zhang Y.-Z., Takamiya S., Capaldi R.A.;
 RT "Tissue-specific differences between heart and liver cytochrome c
 RT oxidase."
 RL Biochemistry 27:4909-4914(1988).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=96216288; PubMed=8638158;
 RA Tsubihara T., Aoyama H., Yamashita E., Tomizaki T., Yamaguchi H.,
 RA Shinzawa-Itoh K., Nakashima R., Yeono R., Yoshikawa S.;
 RT "The whole structure of the 13-subunit oxidized cytochrome c oxidase
 RT at 2.8 A.".
 RL Science 272:1136-1144(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RC TISSUE=Heart;
 RX MEDLINE=99190827; PubMed=10089392;
 RA Tomizaki T., Yamashita E., Yamaguchi H., Aoyama H., Tsubihara T.,
 RA Shinzawa-Itoh K., Nakashima R., Yeono R., Yoshikawa S.;
 RT "Structure analysis of bovine heart cytochrome c oxidase at 2.8 A
 RT resolution."
 RL Acta Crystallogr. D 55:31-45(1999).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
 RC TISSUE=Heart;

```

RX MEDLINE=20235645; PubMed=10771420;
RA Fel M.J., Yamashita E., Inoue N., Yao M., Yamaguchi H., Tsukihara T.,
RA Shinzawa-Itch K., Nakashima R., Yoshikawa S.;
RT "X-ray structure of azide-bound fully oxidized cytochrome c oxidase
RT from bovine heart at 2.9 Å resolution.";
RL Acta Crystallogr D 56:529-535(2000).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC MITOCHONDRIAL ELECTRON TRANSPORT.
CC -1- FUNCTION: THIS PROTEIN MAY BE ONE OF THE HEME-BINDING SUBUNITS OF
CC THE OXIDASE.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15112; CAA33211.1; -.
DR PIR: A00497; OGB07.
DR PIR: I29968; I29968.
DR PIR: S05432; S05432.
DR PDB: 1OCC; 07-DEC-96.
DR PDB: 2OCC; 13-JAN-99.
DR PDB: 1OCR; 29-JUL-99.
DR PDB: 1OCO; 22-JUL-99.
DR PDB: 1OCZ; 22-JUL-99.
DR InterPro: IPR003213; Cyt_c-ox6B.
DR Pfam: PF02297; COX6B; 1.
KW Oxidoreductase; Mitochondrion; Acetylation; 3D-structure.
FT INT_MET 0
FT MOD_RES 1
FT SEQUENCE 85 AA; 10025 MW; CC0E759CC764040A CRC64;

Query Match 66.7%; Score 4; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RCEK 6
DB 38 RCEK 41

RESULT 10
COXG_MOUSE STANDARD; PRT; 85 AA.
AC P56391;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE VIIB (EC 1.9.3.1) (AED).
GN COX6B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Maris M., Hillier L., Allen M., Bowles M., Dietrich N., Dubugue T.,
RA Gaillet S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
RA Scheibler K., Stepien M., Tan F., Underwood R., Moore B.,
RA Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC MITOCHONDRIAL ELECTRON TRANSPORT.
CC -1- FUNCTION: THIS PROTEIN MAY BE ONE OF THE HEME-BINDING SUBUNITS OF
CC THE OXIDASE.

```

```

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: W67100; -; NOT_ANNOTATED_CDS.
DR EMBL: W54267; -; NOT_ANNOTATED_CDS.
DR EMBL: AA500654; -; NOT_ANNOTATED_CDS.
DR EMBL: AA592414; -; NOT_ANNOTATED_CDS.
DR EMBL: W64838; -; NOT_ANNOTATED_CDS.
DR EMBL: AA032941; -; NOT_ANNOTATED_CDS.
DR EMBL: AA139486; -; NOT_ANNOTATED_CDS.
DR MGD: MGI:107460; Cox6b.
DR InterPro: IPR003213; Cyt_c-ox6B.
DR Pfam: PF02297; COX6B; 1.
KW Oxidoreductase; Mitochondrion.
FT INT_MET 0
FT MOD_RES 0
FT SEQUENCE 85 AA; 9940 MW; C674240BCC499F1B CRC64;

Query Match 66.7%; Score 4; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RCEK 6
DB 38 RCEK 41

RESULT 11
RPOL_METHH STANDARD; PRT; 85 AA.
ID RPOL_METHH
AC Q27372;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT L (EC 2.7.7.6).
GN RPOL OR MTH1317.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-DELTA H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashtirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA Modougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nollig J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPOL / EUKARYOTIC RPB1/
CC RPB19 RNA POLYMERASE SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000895; AAB5795.1; -
 DR InterPro: IPR001306; RNA_pol.L.
 DR Pfam: PF01193; RNA_pol.L: 1.
 DR ProDom: PD004240; RNA_pol.L: 1.
 DR PROSITE: PS01154; RNA_POL_L_13KD: 1.
 DR Trnstrfase; Transcription; DNA-directed RNA polymerase;
 KW Complete proteome.
 SO SEQUENCE 85 AA: 9732 MW: CDB234326B4C765A CRC64;

Query Match 66.7%; Score 4; DB 1; Length 85;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRC 4
 DB 72 RDRC 75

RESULT 12
 Y172_BURCE STANDARD: PRT; 94 AA.
 AC P24578;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DE 01-JUN-1994 (Rel. 29, Last annotation update)
 DE INSERTION ELEMENT IS407 HYPOTHETICAL 9.9 KDA PROTEIN (ORF2).
 OS Burkholderia cepacia (Pseudomonas cepacia).
 CC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
 OC Burkholderia.
 RX NCBI_TaxID=292;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 17616 / 249;
 RC MEDLINE=92039004; PubMed=1718819;
 RA Wood M.S., Byrne A., Lesie T.G.;
 RT "IS406 and IS407, two gene-activating insertion sequences for
 RT Pseudomonas cepacia";
 RL Gene 105:101-105(1991).
 CC -1- SIMILARITY: TO THE CORRESPONDING ORF OF X.CAMPESTRIS IS476 AND
 CC OF CLASS-IV STRAINS OF RHIZOBIUM ISR1.
 CC -1- SIMILARITY: IS407, IS476, AND ISR1 COMPRISE A SUBCLASS WITHIN THE
 CC IS3 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL: M82980; AAA25041.1; -
 CC PIR: S28802; S28802.
 CC KW Hypothetical protein; Transposable element.
 CC SEQUENCE 94 AA: 9897 MW: 00FC2E9FB0DA5A2 CRC64;

Query Match 66.7%; Score 4; DB 1; Length 94;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 DB 79 RCEK 82

RESULT 13

MIH_METEN STANDARD: PRT; 105 AA.
 AC O76534;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PROBABLE MOLT-INHIBITING HORMONE PRECURSOR (MEE-MIH).
 OS Metapenaeus ensis (Greasyback shrimp) (Sand shrimp).
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 CC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
 CC Penaeidae; Metapenaeus.
 OX NCBI_TaxID=32278;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eyestalk;
 RX MEDLINE=98368961; PubMed=9701616;
 RA Chan S.-M., Gu P.-L.;
 RT "Cloning of a cDNA encoding a putative molt-inhibiting hormone from
 RT the eyestalk of the sand shrimp Metapenaeus ensis.";
 RL Mol. Mar. Biol. Biotechnol. 7:214-220(1998)
 CC -1- FUNCTION: INHIBITS Y-ORGANS WHERE MOLTING HORMONE (ECDYSTEROID)
 CC IS SECRETED. A MOLTING CYCLE IS INITIATED WHEN MIH SECRETION
 CC DIMINISHES OR STOPS (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE POSTMOLT, INTERMOLT, AND
 CC PREMOLT STAGES OF THE SHRIMP EYESTALKS AND THE BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD CHH/MIH/GIH/VIH FAMILY OF
 CC HORMONES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL: AF076276; AAC27452.1; -
 CC InterPro: IPR001166; CHH_MIH_GIH.
 CC Pfam: PF01147; Crust_neurohorm; 1.
 CC PRINTS: PR00549; HYPRGLYCEMC2.
 CC DR PRINTS: PR00550; HYPRGLYCEMIC.
 CC DR PROSITE: PS01250; CHH_MIH_GIH; 1.
 CC KW Neuropeptide; Hormone; Signal.
 CC FT SIGNAL 1
 CC FT PEPTIDE 29 105 BY SIMILARITY.
 CC FT DISULFID 35 72 BY SIMILARITY.
 CC FT DISULFID 52 68 BY SIMILARITY.
 CC FT DISULFID 55 81 BY SIMILARITY.
 CC SEQUENCE 105 AA: 12117 MW: 6931D21C8F707402 CRC64;

Query Match 66.7%; Score 4; DB 1; Length 105;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRC 4
 DB 69 RDRC 72

RESULT 14
 TRP5_BOVIN STANDARD: PRT; 109 AA.
 AC Q9MYV9;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SHORT TRANSIENT RECEPTOR POTENTIAL CHANNEL 5 (TRPC5) (FRAGMENT).
 GN TRPC5 OR TRP5.
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.

```

OX NCBI_TaxID=9913;
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20379038; PubMed=10816590;
RA Philipp S., Trost C., Warnat J., Himmerkus N.,
RA Schroth G., Kretz O., Nastainczyk W., Cavalie A., Hoch M.,
RA Flockerzi V.;
RT "rpp4 (CCP1) protein is part of native Calcium release-activated Ca2+-
RT like channels in adrenal cells.";
RL J. Biol. Chem. 275:23965-23972(2000).
CC -1- FUNCTION: THOUGHT TO FORM A RECEPTOR-ACTIVATED NON-SELECTIVE
CC CALCIUM PERMEANT CATION CHANNEL. PROBABLY IS OPERATED BY A
CC PROSPATIDYLINOSITOL SECOND MESSENGER SYSTEM ACTIVATED BY RECEPTOR
CC TYROSINE KINASES OR G-PROTEIN COUPLED RECEPTORS. HAS ALSO BEEN
CC SHOWN TO BE CALCIUM-SELECTIVE. MAY ALSO BE ACTIVATED BY
CC INTRACELLULAR CALCIUM STORE DEPLETION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TRANSIENT RECEPTOR FAMILY. STRPC
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ271070; CAC01678.1; -
DR InterPro: IPR002111; Cal_channel_TrpL.
DR InterPro: IPR002153; Trans_recep.
DR Pfam: PF02164; Trans_recep. 1.
KW Ionic channel; Transmembrane; Ion transport; Calcium channel.
FT NON_TER 1 21
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12607 MW; E0068E440512BAC9 CRC64;

Query Match
Best Local Similarity 66.7%; Score 4; DB 1; Length 109;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RCEK 6
Db 36 RCEK 39

RESULT 15
VG33_BPT4 STANDARD; PRT; 112 AA.
AC P13338;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE RNA POLYMERASE-ASSOCIATED PROTEIN GP33.
GN 33.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
OX [1]
RP SEQUENCE FROM N.A.
RP STRAIN=BK536;
RX MEDLINE=89386003; PubMed=2674900;
RA Hahn S., Rueger W.;
RT "Organization of the bacteriophage T4 genome between map positions
RL 150,745 and 145,824.";
RL Nucleic Acids Res. 17:6729-6729(1989).
RN [2]
RP CHARACTERIZATION.

```

```

RX MEDLINE=94148776; PubMed=8106327;
RA Winkelman J.W., Kassavetis G.A., Geiduschek E.P.;
RT "Molecular genetic analysis of a prokaryotic transcriptional
RT coactivator: functional domains of the bacteriophage T4 gene 33
RT protein.";
RL J. Bacteriol. 176:1164-1171(1994).
CC -1- FUNCTION: BINDS TO E. COLI RNA POLYMERASE AND IS REQUIRED FOR
CC TRANSCRIPTION OF LATE GENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15818; CAA3814.1; -
DR PIR: S05558; GMBPT4.
SQ SEQUENCE 112 AA; 12830 MW; 665443899373E04A CRC64;

Query Match
Best Local Similarity 66.7%; Score 4; DB 1; Length 112;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RCEK 6
Db 101 RCEK 104

```

Search completed: February 26, 2002, 01:53:53
Job time: 422 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 01:44:11 ; Search time 45.36 Seconds
(without alignments)
10.076 Million cell updates/sec

Title: US-09-819-136-2_COPY_117_122

Perfect score: 6

Sequence: 1 RDRCEK 6

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83.3	63	2	A70806	probable fdxd prot
2	83.3	110	2	A55991	effector cell prot
3	83.3	183	2	S25465	lipocalin - giant
4	83.3	264	2	T16271	hypothetical prote
5	83.3	337	2	A53041	effector cell prot
6	83.3	555	2	C96667	unknown protein, 7
7	83.3	711	2	T27358	hypothetical prote
8	83.3	1041	2	T29010	hypothetical prote
9	83.3	2139	2	A35672	crumbs protein - f
10	66.7	42	2	T07550	hypothetical prote
11	66.7	51	2	T07563	hypothetical prote
12	66.7	53	2	S23202	hypothetical prote
13	66.7	56	2	U01698	keppa-casein - bov
14	66.7	65	2	S07451	hypothetical 6k pr
15	66.7	62	2	S03858	proteinase inhibit
16	66.7	72	2	T37445	carboxypeptidase A
17	66.7	77	2	H42526	proteolytic 8.3k prot
18	66.7	83	2	S72989	B9R protein - vacc
19	66.7	85	2	A69042	hypothetical prote
20	66.7	86	1	OGBO7	DNA-dependent RNA
21	66.7	88	2	S09416	cytochrome-c oxida
22	66.7	89	2	H69137	proteinase inhibit
23	66.7	90	2	A69163	DNA-dependent RNA
24	66.7	94	2	S28802	hypothetical prote
25	66.7	97	2	T35968	hypothetical prote
26	66.7	100	2	T69107	small conserved hy
27	66.7	108	2	T21795	hypothetical prote
28	66.7	112	1	GMBP4	hypothetical prote
29	66.7	114	1	S48552	gene 33 protein -
					hypothetical prote

30	66.7	118	2	D64372	hypothetical prote
31	66.7	118	2	T49168	hypothetical prote
32	66.7	126	2	F81143	hypothetical prote
33	66.7	127	2	A49269	phospholipase A2 h
34	66.7	128	2	S10046	lysosome (EC 3.2.1
35	66.7	130	2	S10047	lysosome (EC 3.2.1
36	66.7	138	2	S58128	lysosome (EC 3.2.1
37	66.7	139	2	H82521	trp-like protein -
38	66.7	143	2	JC1302	hypothetical prote
39	66.7	144	2	T51047	alpha-lactalbumin
40	66.7	145	2	S13376	lysosome (EC 3.2.1
41	66.7	145	2	S10027	CM2 protein - duru
42	66.7	147	2	H64385	alpha-amylose inh
43	66.7	148	2	T31208	hypothetical prote
44	66.7	152	2	T51796	hypothetical prote
45	66.7	154	2	H70681	hypothetical prote

ALIGNMENTS

RESULT 1
A70806
probable fdxd protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Mar-2000
C:Accession: A70806
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; MUID: 98295987
A: Accession: A70806
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-63 <COL>
A: Cross-references: GB: A1022022; GB: A1123456; NID: g3261554; PIDN: CMA17740.1; PID: e125
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: fdxd
C: Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S]
F: 3-62/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 83.3%; Score 5; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRCE 5
DB 7 RDRCE 11

RESULT 2
A55991
effector cell proteinase receptor 1 splice form 1b - human
C: Species: Homo sapiens (man)
C: Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 19-Jan-1996
C: Accession: A55991
R: Altmeyer, D.C.
Biochemistry 33, 13848-13855, 1994
A: Title: Splicing of effector cell proteinase receptor-1 mRNA is modulated by an unusua
A: Reference number: A55991; MUID: 95034823
A: Accession: A55991
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-110 <ALT>
A: Cross-references: GB: L32866
C: Keywords: alternative splicing

Query Match 83.3%; Score 5; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DRCE 5
 |||||
 Db 58 DRCE 62

RESULT 3
 S25465
 lipocalin - giant toad
 C:Species: Bufo marinus (giant toad)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
 C:Accession: A44456; I50117; S25465
 R:Archer, M.G.; Harms, P.J.; Thomas, T.; Richardson, S.J.; Wettenhall, R.E.; Schneider, G.
 J. Biol. Chem. 267, 23170-23174, 1992
 A:Title: Protein synthesis at the blood-brain barrier. The major protein secreted by am
 A:Reference number: A44456; M0ID:9305464
 A:Accession: A44456
 A:Status: preliminary
 A:Molecule type: mRNA; protein
 A:Residues: 1-183 <AC2>
 A:Cross-references: GB:106806; NID:g211032; PIDN:AAA48554.1; PID:g211033; EMBL:X67952
 A:Experimental source: choroid plexus
 A:Note: Sequence extracted from NCBI backbone (NCBIP:118239)
 C:Superfamily: lipocalin; lipocalin homology
 F:28-179/Domain: lipocalin homology <LTP>

Query Match 83.3%; Score 5; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRCE 6
 |||||
 Db 81 DRCE 85

RESULT 4
 T16271
 hypothetical protein F35D2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16271
 R:Connell, M.
 submitted to the EMBL Data Library, June 1995
 A:Description: The sequence of C. elegans cosmid F35D2.
 A:Reference number: Z18488
 A:Accession: T16271
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-264 <CON>
 A:Cross-references: EMBL:U28741; NID:g861290; PID:g861291; PIDN:AAA68325.1; CESP:F35D2.3
 C:Genetics:
 A:Experimental source: strain Bristol N2
 A:Gene: CESP:F35D2.3
 A:introns: 40/3; 71/3; 160/3; 197/3

Query Match 83.3%; Score 5; DB 2; Length 264;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRCE 6
 |||||
 Db 83 DRCE 87

RESULT 5
 A53041
 effector cell proteinase receptor 1 - human
 C:Species: Homo sapiens (man)

C:Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 21-Jul-2000
 C:Accession: A53041
 R:Altieri, D.C.
 J. Biol. Chem. 269, 3139-3142, 1994
 A:Title: Molecular cloning of effector cell protease receptor-1, a novel cell surface
 A:Reference number: A53041; M0ID:9418797
 A:Accession: A53041
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-337 <ALT>
 A:Cross-references: GB:126245; NID:g456089; PIDN:AAA19687.1; PID:g456090

Query Match 83.3%; Score 5; DB 2; Length 337;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DRCE 5
 |||||
 Db 58 DRCE 62

RESULT 6
 C96667
 unknown protein, 71502-69704 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C96667
 R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marcia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; M0ID:21016719
 A:Accession: C96667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <STO>
 A:Cross-references: GB:AE005173; NID:g10645401; PIDN:AG21519.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F15H21.4
 A:Map position: 1

Query Match 83.3%; Score 5; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DRCE 5
 |||||
 Db 530 DRCE 534

RESULT 7
 T27358
 hypothetical protein Y70G10A.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27358
 R:Lloyd, C.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: Z20354
 A:Accession: T27358
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-711 <WII>
 A:Cross-references: EMBL:AL032660; PIDN:CAA21752.1; GSPDB:GN00021; CESP:Y70G10A.2
 A:Experimental source: clone Y70G10A

C:Genetics:
 A:Gene: CESP:Y70G10A.2
 A:Map position: 3
 A:Introns: 38/1; 157/3; 246/3; 297/1; 404/2; 467/2; 490/3; 599/2; 679/3

Query Match 83.3%; Score 5; DB 2; Length 711;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCEK 6
 DB 197 DRCEK 201

RESULT 8
 T29010
 hypothetical protein ZK328.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29010
 R:Favelli, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of C. elegans cosmid ZK328.
 A:Reference number: Z20552
 A:Accession: T29010
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1041 <FAV>
 A:Cross-references: EMBL:U50193; PIDN:AA91251.1; CESP:ZK328.1
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:ZK328.1
 A:Introns: 25/1; 64/1; 143/1; 231/2; 285/3; 487/1; 567/2; 643/1; 937/3; 987/3; 1018/3

Query Match 83.3%; Score 5; DB 2; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCEK 6
 DB 916 DRCEK 920

RESULT 9
 A55672
 crumbs protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 11-Jan-2000
 C:Accession: A55672
 R:Teppas, U.; Theres, C.; Knust, E.
 Cell 61, 787-799, 1990
 A>Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila
 A:Reference number: A55672; MUID:90263104
 A:Accession: A55672
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2139 <TEP>
 A:Cross-references: GB:M33753
 A>Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue
 C:Genetics:
 A:Gene: FlyBase:ctb
 A:Cross-references: FlyBase:FBgn0000368
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: transmembrane protein
 F:352-385/Domain: EGF homology <EGX1>
 F:392-424/Domain: EGF homology <EGX1>
 F:691-722/Domain: EGF homology <EGF>
 F:767-799/Domain: EGF homology <EGF3>
 F:1878-1914/Domain: EGF homology <EGX2>

Query Match 83.3%; Score 5; DB 2; Length 2139;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DRCEK 6
 DB 578 DRCEK 582

RESULT 10
 T07550
 hypothetical protein 42d - Japanese black pine chloroplast
 C:Species: chloroplast Pinus thunbergiana (Japanese black pine)
 C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
 C:Accession: T07550
 R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Suglura, M.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
 A>Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen
 A:Reference number: Z16030; MUID:95024047
 A:Accession: T07550
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-42 <NAK>
 A:Cross-references: EMBL:DJ7510; NID:g529643; PIDN:BA04426.1; PID:g1262711
 C:Genetics:
 A:Genome: chloroplast
 C:Keywords: chloroplast

Query Match 66.7%; Score 4; DB 2; Length 42;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 DB 16 RCEK 19

RESULT 11
 T07563
 hypothetical protein 51b - Japanese black pine chloroplast
 C:Species: chloroplast Pinus thunbergiana (Japanese black pine)
 C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
 C:Accession: T07563
 R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Suglura, M.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
 A>Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen
 A:Reference number: Z16030; MUID:95024047
 A:Accession: T07563
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-51 <NAK>
 A:Cross-references: EMBL:DJ7510; NID:g529643; PIDN:BA04439.1; PID:g1262724
 C:Genetics:
 A:Genome: chloroplast
 C:Keywords: chloroplast

Query Match 66.7%; Score 4; DB 2; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRC 4
 DB 20 RDRC 23

RESULT 12
 S23202
 kappa-casein - bovine
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
 C:Accession: S23202

R.Rasmussen, L.K.; Hojrup, P.; Petersen, T.E.
 Eur. J. Biochem. 207, 215-222, 1992
 A:Title: The multimeric structure and disulfide-bonding pattern of bovine kappa-casein.
 A:Reference number: S23202; MUID:92331656
 A:Accession: S23202
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-53 <RAS>

Query Match 66.7%; Score 4; DB 2; Length 53;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 ||||
 Db 10 RCEK 13

RESULT 13
 JQ1698
 hypothetical 6K protein - Plectonema sp. plasmid pRF1
 N:Alternate names: ORPA protein
 C:Species: Plectonema sp.
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
 C:Accession: JQ1698
 R;Perkins, D.R.; Barnum, S.R.
 Plasmid 28, 170-176, 1992
 A:Title: DNA sequence and analysis of a cryptic 4.2-kb plasmid from the filamentous cyanobacterium Plectonema boryanum
 A:Reference number: JQ1698; MUID:93028891
 A:Accession: JQ1698
 A:Molecule type: DNA
 A:Residues: 1-56 <PER>
 A:Experimental source: strain PCC6402
 C:Genetics:
 A:Genome: plasmid

Query Match 66.7%; Score 4; DB 2; Length 56;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RDRC 4
 ||||
 Db 23 RDRC 26

RESULT 14
 S07451
 proteinase inhibitor 5.II - snake-locks sea anemone
 C:Species: Anemonia sulcata (snake-locks sea anemone)
 C>Date: 31-Dec-1990 #sequence_revision 09-May-1997 #text_change 09-May-1997
 C:Accession: S07451; B27222
 R;Wunderer, G.; Machleidt, W.; Fritzt, H.
 Meth. Enzymol. 80, 816-820, 1981
 A:Title: The broad-specificity proteinase inhibitor 5.II from the sea anemone Anemonia sulcata
 A:Reference number: S07451
 A:Accession: S07451
 A:Molecule type: protein
 A:Residues: 1-59 <WUN>
 A>Note: 13-Arg, 16-Gly, 17-Gly, 25-Leu, 28-Arg, and 39-Arg were also found
 R;Krebs, H.C.; Habermehl, G.G.
 Naturwissenschaften 74, 395-396, 1987
 A:Title: Isolierung und strukturaufklärung eines haemolytisch aktiven peptids aus der Seeanemonie Anemonia sulcata
 A:Reference number: A94700
 A:Accession: B27222
 A:Molecule type: protein
 A:Residues: 1-38, 'R', '40', 'B', '42', 'BB', '45-48', 'ZZ', '51', 'Z', '53-62 <KRE>
 C:Superfamily: basic proteinase inhibitor; animal kunitz-type proteinase inhibitor homol
 C:Keywords: serine proteinase inhibitor
 F:5-55/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match 66.7%; Score 4; DB 2; Length 62;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 ||||
 Db 29 RCEK 32

RESULT 15
 S03858
 carboxypeptidase A inhibitor - pig roundworm
 C:Species: Ascaris suum (pig roundworm)
 C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 18-Jun-1993
 C:Accession: S03858
 R;Homandberg, G.A.; Litwiler, R.D.; Peanasky, R.J.
 Arch. Biochem. Biophys. 270, 153-161, 1989
 A:Title: Carboxypeptidase inhibitors from Ascaris suum: the primary structure.
 A:Reference number: S03858; MUID:89192365
 A:Accession: S03858
 A:Molecule type: protein
 A:Residues: 1-65 <HOM>

Query Match 66.7%; Score 4; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RCEK 6
 ||||
 Db 33 RCEK 36

Search completed: February 26, 2002, 01:51:44
 Job time: 453 sec

This Page Blank (uspto)